

Molecular biology: suffering from shock

THE recent meeting of molecular biologists convened by COGENE, the Committee on Genetic Experimentation of the International Council of Scientific Unions, was an instructive affair—but not in the way the organisers intended. The lesson came in their treatment of the press. First the meeting was open; then it was closed to journalists. Then it was open to three journalists picked by the Association of British Science Writers, but the proceedings were to be 'off the record'. Then at the last moment all restrictions were lifted.

To be fair, the organisers of the meeting—COGENE and the Royal Society—admitted they had made a mistake; they were under pressure, it seemed, from biologists who had had bad experiences with the US press. And the press is certainly no angel. But the meeting's attitude to it was symptomatic of a deeper malaise: that despite the promise of the 'Berg letter' (*Genesis*, someone called it) which launched molecular biology into politics, biologists have failed to grasp the nettle of their relationship to the public. The press, after all, represents all the fears and frustrations of its readers: it is the public writ in 48 point bold caps.

However the molecular biologists—if the COGENE meeting was a guide—now feel that their subject has no right to be in the public arena at all. The Berg letter called for a pause to assess the risks of the recombinant DNA technique, we have had the pause, and the risks are found to be negligible. So why don't the NIH, and GMAG, and the other regulatory agencies just go away? That was the strong mood of the meeting. "We want to achieve re-entry," someone said. But as Sir Frederick Warner, one of the assessors at the Windscale inquiry into nuclear reprocessing, told the meeting, the three great fears have been raised in the public mind: carcinogenesis, mutagenesis, and teratogenesis. The public won't go away. Re-entry into the womb is no longer possible, even if it seems rational.

Paul Berg's two revolutionary impacts on molecular biology—the recombinant DNA technique itself, and (with the other 10) the Berg letter—have left molecular biologists in a state of shock. Within six years they have found an immensely promising research technique; they have entered biotechnology, with undreamt-of profits to be made; and they have entered politics with their perfectly rational whistle-blowing over conjectural risks. All this happened at just the time when a new breed of radical scientists was searching for something to cut its teeth on, and when the regulation of science and technology was becoming fashionable among governments. So no wonder the biologists are reeling.

No doubt the organisers of the COGENE meeting intended that it should come to terms with this new climate.

But it only reacted to it. Four major issues facing molecular biology were either not addressed or left unresolved. First, what really are the arguments that the risks are negligible? Although there were interesting talks—such as that of Walter Bodmer on species barriers—which were relevant to the question, there was no scientific *debate* of the matter. Where were the critics—and there are scientific critics, with genuine arguments—to bring matters to a head? To the outside observer the view that risks are negligible went through on the nod. If the biologists are to face the public they must do better than communicate a mood; they must construct arguments that are for more than internal consumption. To do that they must embrace their heretics, not spit at them.

Second, there was not a minute of discussion of the new procedure for risk assessment devised by Sidney Brenner and recently adopted by Britain's Genetic Manipulation Advisory Group. Confounding its critics who complain that biological knowledge is too patchy for it to work, this method is already in use in GMAG, and has resulted in the down-grading of a number of experiments. If it is false, it should be exposed; if it is correct, being simple and logical it could be the basis of communicating the biologists' new mood to the public.

Third, if there was no scientific debate, neither was there political debate, despite the fact that it is clear that the biologists' re-entry must be a political one. The most moderate voices, such as that of Mark Richmond, showing sensitivity to politics and public opinion, sounded radical against the growls of the rednecks, and were submerged.

Fourth and last, the role of molecular biology in technology, where recombinant DNA is the microprocessor chip of the pharmaceutical industry, was dealt with only at the level of bioengineering and gee-whizzery. And yet here is a technique which has already generated a new social world for the biologist, who is becoming suspicious of his colleague's motives (does he want my enzyme for profit or for science?) and hungry for a new kind of success. Is that not a significant development? Why was it not a subject for debate? Further, what impact will this technique have on medicine and agriculture? Should it be left to the few secretive new firms that are engaged on it, or do we have a chance to create a humane and democratic technology? These questions were ignored.

Molecular biologists are firmly in politics now, and it is totally irrational to ignore the fact. Nuclear physicists should envy them: for the physicists generated their technology first, became committed to it, and burnt their political fingers later. The biologists have the opportunity to learn their politics first, and grow with it. They should not miss the opportunity. □



California calls for nuclear shutdown

David Dickson reports from California on reactions to the accident at the Harrisburg nuclear power plant

THREE years ago, the citizens of California voted by a two to one majority against any curb on the development of nuclear power. Last week, following the accident at the Three Mile Island nuclear power plant in Harrisburg, Pennsylvania, which took over a week to bring under control, and resulted in the discharge of radioactive isotopes into the local environment, an informal survey by a San Francisco newspaper based on telephone calls from readers revealed a 56% majority in favour of shutting down all nuclear power plants.

No one pretends that the California figures are a true reflection of attitudes across the US. California's maverick tradition on "lifestyle" issues, together with a high incidence of sunshine that makes it relatively well placed with respect to alternative sources of energy, both tend to inflate public opposition to nuclear power.

However, the shift in attitude, to judge from press reports, does reflect a public mood. Governor Jerry Brown, his presidential ambitions not far below the surface, has argued with both the Nuclear Regulatory Commission and the local utility company for a temporary shutdown of the Rancho Seco power plant in California (which like the Harrisburg plant was designed by the engineering firm Babcock and Wilcox). And even some of those who had previously been staunch supporters of the nuclear industry are voicing concern that perhaps the Harrisburg incident indicates that not all is as it should—or had been made out—to be.

Defenders of the industry are using the fact that the amount of radiation released to the atmosphere has been minimal, and well below levels claimed to have a harmful effect on humans, to argue that the industry had shown that its safety precautions worked. "Reactors are designed to survive the failure of instruments, which is why they have back up systems," said Dr Edward Teller, so-called "father" of the hydrogen bomb, and now with Stanford University's Hoover Institution. "The worst-case estimate is that some people were exposed to about 50 millirems. Airline hostesses get that in about 50 hours of flight, which they can probably accumulate in about three weeks," Dr Teller said.

But Professor Joseph Rotblat, Emeritus Professor of Physics at St. Bartholomew's Hospital, London University, points out that if the official figure of 25 millirem per hour within a five-mile radius was sustained for the first week of the accident, the 24,000

inhabitants within the five-mile radius could have received a total of 5 rems. This exceeds the maximum allowable dosage for radiation workers of 3 rems in any 13-week period and is ten times the permissible value for the population at large. Official estimates attribute one cancer case per 10,000 population per rem but, says Rotblat, "this estimate is perhaps a factor of ten too small", which would imply an additional 120 cancer cases because of the accident.

If the incident did little to shake the support of firm supporters of nuclear power, neither did it come as much surprised to nuclear critics. "This is just the type of thing that we have been warning people would happen," said Mr Bob Judd, director of California's Office of Appropriate Technology in Sacramento.



"But we've put a million dollars into Meltdown at Rancho Seco!"

Where there has been a perceptible shift, however, is in those who have yet to come down firmly on one side or another. In the California state legislature, a number of politicians who still support the development of nuclear energy voiced concern at the implications of Harrisburg—for example, at the fact that the formation of the hydrogen bubble which formed at the top of the containment vessel, and gave the greatest source of concern, had never entered anyone's calculations as a potential event—and supported Governor Brown's call for a temporary closure of Rancho Seco.

Similarly on university campuses some of those previously sympathetic to the anti-nuclear movement, but sceptical of its apparent "anti-technology" stance, admitted that the Harrisburg incident had shifted their perspective. "I am beginning to believe that there is something inherently unstable about nuclear power; that unlike other energy sources, the more you develop it, the more you have to develop systems to keep it under control," says Professor Charles Schwarz, professor of physics at the University of California at Berkeley.

The incident hung like a black cloud

over the proceedings of a San Francisco conference on "science, technology and the human prospect" which had been organised jointly by the Electric Power Research Institute and the Thomas Alva Edison Foundation as part of a year-long commemoration of the invention of the electric light one hundred years ago.

The opening day of the conference coincided with Governor Brown's call for the closing of Rancho Seco. Dr Chauncey Starr, Vice-President of FPRI, admitted at a press conference that he would have liked to discuss some of the broader issues being raised by the conference, billed as "a unique opportunity to explore the impact of science and technology on society's development". But all the questions, inevitably, turned on Harrisburg.

"Give us the facts. We want facts, not opinions," demanded one journalist, as Dr Starr explained that at present the precise sequence of events within the reactor that had led to the accident was unknown, but assured people that the levels of radiation to which the local community had been exposed were "so low that they were below any physiological danger to anyone, including pregnant women."

The accident did not change the long-term picture, he said. The initiating circumstances—the apparent failure of a pump circulating the cooling water—was not uncommon, and the statistical probability of the accident fell within the range that had been predicted for the nuclear power programme over 400 reactor years of operation. "Risk analysis is not going to be changed in any particular way by this," Dr Starr said.

"Although something happened in this plant that we do not yet fully understand, no one has been hurt, and there is no public danger that any of us can see. I do not think that shutting down nuclear power plants is a way to increase public confidence. We will continue to do our best to get the public to understand that there are sufficient assets to nuclear power compared to other energy sources, and that the health risks remain very much less. Remember that even steam boilers blow up."

Back in the less charged atmosphere of the conference, speakers talked in general terms about the need to maintain confidence, that society could learn to control the technology that it had developed, and claimed that the Harrisburg incident should reinforce—rather than dent—this confidence (though most were reluctant to get into a detailed discussion of the incident itself). □



President Carter is briefed in the control room at Three Mile Island by Harold Denton of the NRC (left).

Harrisburg: counting the cost

At Harrisburg, the accident still poses enormous clean-up problems. The radiation level inside the reactor containment dome reached 30,000 rems per hour, and officials face the task of cleaning up 250,000 gallons of radioactive water from inside the reactor vessel itself. "We are looking at plans to bring the reactor to a cold shutdown without an increased leakage from the plant" said Harold Denton, Director of Reactor Regulation of the Nuclear Regulatory Commission (NRC). But Senator Morris Udall said it would be months before any clean-up could begin. "if, indeed, a clean-up is possible".

In the meantime, there have been press reports blaming the Metropolitan Edison Company for violating Federal regulations for reactor operation. The *International Herald Tribune* cites detailed NRC reports that valves controlling the emergency water supply to the reactor's cooling system had been closed for two weeks for routine maintenance. Mr Darrell Eisenhut, an engineer for the NRC, told a press conference that in addition the analyses showed that the main cooling system had been turned off at the wrong time and that four auxiliary pumps were disengaged in violation of Commission regulations. These faults combined to cause 60,000 gallons of radioactive water to flood the reactor chamber to a depth of eight feet. The zirconium in the reactor fuel cladding then combined with the water to release hydrogen gas which created a large 1,000 cubic feet bubble at the top of the vessel, causing the fuel rods to overheat and rupture releasing fission products into the water.

In a letter to *The Guardian* in London, Sir Martin Ryle of the Cavendish Laboratory in Cambridge said that the deficiencies of the emergency core cooling system of

the Harrisburg family of pressurised water reactors (PWRs) had "been known publicly for five years and by the US Atomic Energy Commission for considerably longer". According to Ryle, the system consists of four separate methods of cooling and relies on correct working of monitored temperature, pressure and waterlevel and the correct sequencing of the various pumps and valves. "It is an excessively complicated system that has never been tested except for 'tests' by computer simulation", said Ryle. He added that the highly dangerous hydrogen bubble should have been predicted as a matter of "A-level textbook knowledge" or failing that as the result of an "afternoon's experiment".

In the financial world, the accident touched off heavy sales of nuclear power related stocks. The US stock market as a whole suffered a broad decline which was attributed to investors worrying over the implications of the accident for the US energy situation. However, stock in Columbia Pictures, owners of *The China Syndrome*, a film about a nuclear meltdown "somewhere in Pennsylvania", rose as its box office takings reached a new high. An anti-nuclear group in South Carolina reported that their phones had been busy 16 hours a day. "People call up and say I can't believe it. I saw the movie and then came home and saw the same thing on the 11 o'clock news."

Daniel Ford, executive director of the Union of Concerned Scientists, called on President Carter "to seek immediate removal" of Dr Joseph Hendrie as chairman of the NRC. Ford was reported as saying that Hendrie had participated in a "far reaching cover-up of critical nuclear safety difficulties".

Joe Schwartz

Joe Schwarz monitors world wide reactions

'We all live in Pennsylvania..'

IN THE UK, the Secretary of State for Energy, Tony Benn, said that before the accident he had faced continuous pressure to "go American" and replace Britain's gas cooled reactors with American PWRs similar to the Harrisburg design. Pressure came from advisers, civil servants, and the nuclear industry itself, but he rejected the plan. Benn said the Harrisburg accident proved that "energy decisions must be kept under democratic control of Parliament". Prime Minister James Callaghan told Parliament: "We have been very wise in concentrating on a safer type of reactor." One of the three firms competing for the order of Britain's first PWR is Babcock and Wilcox; the others are the West German company Kraft Werke Union and a consortium which includes Rolls Royce and Northern Engineering.

France and West Germany, the two most committed nuclear nations in Europe, sent scientific teams to Harrisburg to inspect the accident. Two West German experts reported that the reactor was closer to a core meltdown than was publicly admitted.

The French Government announced its intention to proceed as usual with nuclear development, but the Opposition and the trade unions have called for a re-examination. Francois Mitterand, leader of the Socialist Party, accused the government of secrecy and high handedness in pushing through its power programme. The Socialist minority in parliament called for a halt to France's plan to build nine additional PWRs similar to the Harrisburg reactor in the next five years.

Prime Minister Raymond Barre told the European Press Club: "We must multiply our security measures but France cannot renounce nuclear energy". Minister of Industry Andre Giraud said: "An accident of this type has been taken into account in the design of French nuclear plants."

Le Monde reported numerous criticisms of the Government's stance. The Committee of Interregional Ecological Movements (CIME) demanded an immediate halt to the French nuclear programme because of the danger to urban centres. A Harrisburg-type accident, CIME said, would contaminate 300,000 people around the Fessenheim reactor (Mulhouse), 100,000 around Flamanville (Cherbourg) and 500,000 around Pellerin (Nantes).

The most dramatic action in France

was at the Mediterranean naval base near Toulon with the bombing of several million dollars worth of reactor equipment intended for export to Iraq, West Germany and Belgium. Responsibility was claimed by a group of militant ecologists in a call to *Le Monde*. The caller said: "We have succeeded in neutralising machines dangerous to human life." However, French officials are reported as believing that the bombing may have been an Israeli action directed primarily against the Iraqi reactor programme.

In West Germany, 35,000 demonstrators at the Gorleben hearing into nuclear reprocessing plans in Lower Saxony chanted, "we all live in Pennsylvania." Inside the hearings, a French scientist, Yves Lenoir of the mining academy of Fountainebleau, walked out of the meeting of 60 international experts saying that the hearings were a sham and had no influence over the decision. Lenoir said that any discussion of safety that did not take into

account the Harrisburg events was academic.

But the Ministry of the Interior in Bonn said that German safety standards were high compared with international standards but they would be "toughened up" drastically. "Utilities and the nuclear industry will have no chance to ease safety precautions," a spokesman said.

In Sweden the Opposition party demanded that the Ringhals Two plant near Goteberg—a PWR similar to Harrisburg should be closed for inspection; and Premier Ola Ullsten announced a national referendum on the country's nuclear programme. Danish politicians urged greater debate before their country took a decision to build nuclear reactors, and in Belgium the mayor of the town of Huy, 40 miles south-east of Brussels, ordered the closing of the 870 MW reactor, Tihange 1, saying that the Harrisburg accident had shown the town's emergency preparations to be inadequate.

In Japan, Premier Masayoshi Ohira said there would be no change in Japan's programme which currently has 19 reactors producing 11% of the country's energy. But 100 demonstrators staged a sit-in at the Ministry of International Trade and Industry calling for a thorough inspection of all Japan's nuclear power plants. The ministry itself was holding an emergency meeting with officials from 15 areas to discuss nuclear safety measures.

In Eastern Europe, comment on Harrisburg was limited. Soviet television devoted 15 minutes to the event on 2 April saying, "the accident has provoked a profound anxiety and continues to alarm the American people. A particular indignation has been aroused by the fact that the energy monopolies, in searching for profits, do not take the necessary measures in order to assure the safe functioning of nuclear power stations." □

Guidelines should go, DNA meeting concludes

Eleanor Lawrence reports on a meeting where biologists scourged themselves for going public on conjectural risks

LAST week in the village of Wye, deep in the Kent countryside, an audience predominantly composed of molecular biologists overwhelmingly reiterated the now widely-held view that recombinant DNA research poses no special risks.

The meeting was convened by the Royal Society and COGENE, the Committee on Genetic Experimentation of the International Council of Scientific Unions, to discuss the status of recombinant DNA work and the guidelines controlling it.

Essentially most scientists working in the field now believe that the original fears were based on bad scientific judgment, and that recombinant DNA experiments at the very worst can pose no more hazard than that of working with the most dangerous organism involved in the experiment. Therefore, they argue, regulations for recombinant DNA research are unwarranted and should be abolished. The inconsistency inevitable in guidelines designed to guard against conjectural hazards and the bureaucracy involved in their implementation pose a threat to the freedom of scientific enquiry.

The scientific basis for the change of heart appears to rest first on advice from experts in infectious diseases that it is virtually impossible to convert the laboratory strain of the common gut organism *E. coli* into an epidemic pathogen by the random insertion of a block of foreign genes.

The fear that the insertion of animal virus genes into *E. coli* would result

in a new route for the virus to bypass normal host defences is now also held to be groundless. The current conventional wisdom is that work with cloned viral DNA poses, if anything, even less risk than work with the virus itself, and that cloned viral DNA fragments offer the safest way to study the molecular biology of the most lethal viruses such as Lassa or smallpox.

Joe Sambrook of Cold Spring Harbor Laboratory observed that had Professor Bedson been working in Birmingham with cloned smallpox sequences rather than the complete virus, both he and Janet Parker would be alive today. (The World Health Organisation is indeed considering the possibility that cloned fragments of smallpox DNA might be the safest way of conserving the smallpox genome for posterity). Although little direct evidence addressing this question was available in 1977—when the Ascot Workshop's recommendation of such views to the NIH was influential in gaining considerable relaxation of guidelines for work with animal viruses—experimental support in the case of certain DNA viruses, at least, has recently been obtained from the NIH 'worst case' polyoma virus experiment (see box).

Participants in the debate also marshal evolutionary arguments, such as the growing appreciation that genetic exchange occurs across wide species barriers in microorganisms, as general ammunition. But it is not clear that

these arguments necessarily address the particular point that still worries those who see the need for guidelines. They ask to be assured that the specific product of any given recombinant DNA experiment is not going to be hazardous to those who may be exposed to it, either in the laboratory, or in the general environment.

Among those who call for the end of regulation a more subjective attitude is that expressed most forcibly by J. D. Watson of Cold Spring Harbor. Watson now attributes the call for the moratorium as mixture of fears over research with tumour viruses themselves and an attack of mild liberal guilt, and considers that he and this fellow signatories displayed a complete lack of scientific judgment. "We were jackasses," he told the conference. "It was a decision I regret; one that I am intellectually ashamed of". Watson adopted this position soon after the original 'Berg letter' was written.

Another signatory, Stanley Cohen of Stanford University, also felt the group's original action was irresponsible on scientific grounds, as well as politically naive. It was based simply on a "lack of certainty there was no risk" and was therefore an "irresponsible scientific argument".

The third signatory present, Norton Zinder, holds a somewhat different view. Although he now thinks their original fears to be groundless, in the circumstances as they saw them at the time there was no other action that could be taken.

Given that they now largely have the support of their scientific colleagues on

guideline committees, the problem now facing those who wish to get rid of public guidelines was neatly summed up by Mark Richmond, a member of the UK Genetic Manipulation Advisory Group. Publication of the Berg letter, he pointed out, was in effect a political act even though there had been no political intention. But through this political act the subject has lost its innocence, and it is now going to be hard for the scientists involved to regain a position in which their views can be regarded as objective.

"How are you going to reassure the public that you're not arguing from self-interest", said Richmond, "but that you are now arguing objectively—particularly when, looked at from the outside, you seem to make enormous quantum jumps between what you require for conditions at one time and what you required for the same experiments a couple of years ago".

R. Pritchard, Professor of Biochemistry at Leicester and a well-known opponent of the UK GMAG reiterated his opinion that the effect of the Berg letter was largely due to the eminent names appended to it, and that its message could only be countermanded by the authors themselves. They should state publicly and unequivocally what they now believe, he said, to the loudest round of applause of the meeting.

Not all the participants, however, judged the display of public recantation before the Inquisition to be warranted. In the perception of a historian of science, Charles Weiner of MIT, the recombinant DNA debate has not been the wasteful and illborn episode it seems to most of those concerned.

Weiner has followed the tortuous course of the debate from the beginning and has collected in his *Recombinant*

'Worst case' experiments show low risk

THE original fears which led to the call for a moratorium were largely prompted by the prospect that cloning animal viruses, and especially tumour viruses, would provide them with a new or more effective way of overcoming host defences. The first results, reported recently, from the NIH 'polyoma experiment' designed to test this proposition were reassuring. Tests of the infectivity of complete copies of polyoma DNA cloned in plasmids or phage lambda showed it to be non-infectious or much less infectious (by a factor of 10³) than polyoma virus itself.

One of the NIH experimenters, Dr M. Martin of the National Institute of Allergy and Infectious Diseases, told the Wye conference of results they now have on the tumorigenicity of cloned polyoma DNA. *E. coli* carrying plasmids containing one copy of polyoma DNA produced no tumours when introduced into newborn hamsters; polyoma virus introduced in comparable conditions produces tumours in all animals infected.

The circular recombinant plasmid DNA itself produced a few tumours when given in quantities equivalent to 10⁶ times those at which polyoma virus is invariably tumorigenic. As expected from the previous infectivity experiments, bacteriophage lambda-containing dimeric polyoma DNA also caused tumours in a few cases.

The NIH experiment was delayed first by the regulations against the cloning of animal viruses and then by a private legal suit brought against NIH which sought to prevent the experiment. A group of European scientists sponsored by EMBO to undertake similar polyoma infectivity experiments completed initial *in vitro* infectivity tests (which came to essentially the same conclusions as the NIH study) last year under the then less restrictive UK regulations, but have been unable to carry out animal infectivity or tumorigenicity tests because of the lack of any approved facility for such animal experiments in Europe.

In contrast to the US team they were able to obtain dimeric polyoma DNA cloned in plasmids (the NIH scientists could only obtain dimeric polyoma inserts in lambda) which will now be tested for animal infectivity and tumorigenicity at Fort Detrick.

□

DNA Archive at MIT much of the formal documentation, supplemented by the informal accounts of those involved. Although molecular biologists found the experience of coming face to face with the apprehensions of society a traumatic one, they have, in Weiner's opinion, come through with their science and honour relatively unscathed.

Despite the trauma and despite the

bureaucracy said Weiner, recombinant DNA research flourishes. Has the whole debate and its attendant publicity, he wondered, stimulated rather than repressed the research? But he was concerned at the attitude he now sees amongst younger scientists, to whom the message of the recombinant DNA debate now seems to be "shut up or be shut down".

□

US regulators countenance abolishment of the guidelines

INDICATIONS that the administrators most closely concerned are sympathetic to the efforts of the US scientific community to disentangle itself from the coils of recombinant DNA guidelines were provided by the statement at the conference from the Director of NIH, Donald Fredrickson.

Fredrickson's decision to go public on the proposed guidelines early in 1976 exposed the scientists to the full force of a public debate. But Fredrickson maintains that this was the only way to release the rising tension and "to prepare to defend whatever actions would be taken, against certain criticism".

The first NIH guidelines were released in June 1976, and greatly relaxed guidelines came into force early this year. In his Wye address, Fredrickson described the latest guide-

lines as a "new set of rules painfully formulated during this unprecedented curtailment of experimentation in biology".

In his view one of the important achievements of the revision was to provide for continuous and orderly evolution of the rules—"even to their eventual elimination when the need passes". Further revisions are due to come into force within the next few weeks, notably the approval of strains of *Bacillus subtilis* and *Saccharomyces cerevisiae* host-vector systems for various levels of containment, a provision hitherto lacking in the NIH guidelines.

H. DeWitt Stetten, former Chairman of the Director's Recombinant DNA Advisory Committee, used his talk to describe his personal route to disenchantment. After battling with the

myriad inconsistencies embodied in the first guidelines he eventually came to the conclusion that they had "got off the track badly" and were doing things that were very wrong—given that the mission of NIH was to support and encourage research.

He decided that the hazards had been overstated as no one could give him any evidence that there was any hazard. He concluded that "we had erected a bastion against a phantom", and that the barriers built to exclude ghosts were composed of scientists' time, effort and money. He has proposed that the guidelines be reduced to one simple rule: "the appropriate conditions of containment for recombinant DNA experiments should be those of the most virulent micro-organism entering into the experiment".

□

news in brief

Lower Saxony stalls for time in wake of Harrisburg:

The six-day international hearing on the planned nuclear reprocessing facility at Gorleben in West Germany was very much influenced by the events of Harrisburg. Most of the proponents or "neutral participants" were obviously interested in gaining time. The organisers of the hearings—the state government of Lower Saxony—and the increasingly involved public (the largest anti-nuclear demonstration in Germany with over 100,000 participants took place during the week) were presented with two entrenched and irreconcilable camps. The pro-nuclear lobby insisted that the gigantic project could be realised, and proposed 1994 as a new start-up date for the reprocessing plant. The anti-nuclear lobby maintained that the project was more complicated, dangerous and expensive than simply storing nuclear waste underground. The Lower Saxony government is expected to take its time in assessing the meeting and may defer a decision until after elections for the European Parliament and the state government of neighbouring Schleswig-Holstein. Surprisingly the government granted permission for further trial drilling to a depth of 250m but only outside the area of the salt domes where final storage is planned. The main aim of the drilling is to find out about the surrounding hydrological structure.

From Klaus Höpfner in Hanover

US tries to stop proliferation in Pakistan:

The State Department has announced that the US is cutting off development aid to Pakistan amounting to £20m this year and £23m next year in an effort to prevent Pakistan building a uranium enrichment plant. The action was required by a Congressional ruling to deny aid to countries rejecting nuclear inspection or safeguards. US officials became suspicious because Pakistan does not have reactors that can use enriched uranium. According to *The Guardian*, a spokesman from the Pakistani Foreign Ministry in Islamabad "angrily rejected" claims that Pakistan was planning to build an atomic bomb. "Somehow it gives rise to more concern if an Islamic country carries out research into some areas of nuclear technology than if other countries in the world do so" said the spokesman.

US intelligence sources claim that Pakistan has been surreptitiously buying equipment to make a gas centrifuge enrichment facility. The centrifuge was purchased in Holland and high frequency electric power converters were purchased from Emerson Electric in the UK. The UK government stopped the export after it was told about its potential military use.

ESA and NASA agree on joint solar-polar mission:

The European Space Agency and the US National Aeronautics and Space Administration have signed a memorandum of understanding for a joint International Solar Polar Mission (ISPM) to be launched in 1983. ESA and NASA will each provide a spacecraft, and ESA will supply software and support personnel to manage the ESA flight operations and data processing at the mission control and computing facility which will be provided by NASA. The two-spacecraft mission is designed to observe the Sun from above and below its polar regions. The spacecraft will be launched simultaneously by the Space Shuttle and will travel in the ecliptic plane out to Jupiter where they will swing around under the influence of Jupiter's gravity to head back towards the Sun on orbits out of the ecliptic plane, one heading towards the Sun's north pole, the other towards the Sun's south pole. The spacecraft will circle the Sun on this polar orbit and then head back to Jupiter. The period from

launch to the second polar passage will be about five years. The mission will also investigate the interplanetary medium during the initial Earth to Jupiter run and the Jovian magnetosphere during the Jupiter fly by.

Comprehensive schools maintain standards in science:

Pupils at inner London's non-selective (comprehensive) schools obtained almost the same number of O level and CSE grade 1 passes in biology, chemistry and physics as those from selected intake schools in 1977, and the comprehensive schools achieved these results despite having inferior science facilities and spending less time on the subject. These are the conclusions of a report issued by the Inner London Education Authority science inspectorate. The report expressed concern about the imbalance in facilities, and emphasised that the excellent examination results achieved by dedicated comprehensive school staff should "in no way be taken as an excuse for doing nothing to improve the resources in these schools".

The poorest science facilities were in non-selective mixed and girls schools. According to the report, the main reason for the differences was probably the greater range of subjects taught in comprehensive schools. They had proportionately less resources to spend on science subjects than selective schools since they had to teach subjects such as home economics, design and technology. Selective schools tend to offer only traditional subjects. However improvements are in progress.

Dr John Price, ILEA senior science inspector, said the authority's schools subcommittee had recommended setting up a central task force of 30 science teachers by 1980-81 who will help out in any school where there is a temporary shortage of science staff. The schools subcommittee has also pressed for the urgent review of the staffing and pay points system which causes comprehensive schools to have fewer experienced teachers. Finally, all schools will be given minimum laboratory facilities which will be improved in later years when more money is available.

India prepares for monsoon measurements:

India is making major preparations for the forthcoming Monsoon Experiment (MONEX-79) which will enhance understanding of monsoon circulation around the three seas bordering India: the Arabian Sea, the equatorial Indian Ocean, and the Bay of Bengal. Four ships and one plane from India will be supplemented by five Soviet ships and three US planes with further information to be gathered by meteorological balloons and cyclone warning radar stations situated along the Indian coast. Marine weather reports from merchant ships and upper air reports from commercial aircraft will also be analysed.

Three of the Indian ships are expected to be equipped with Navaid sounding systems obtained through the World Meteorological Organisation and the fourth with an Indian-made Omega system to enable them, for the first time, to make systematic recordings of upper winds over the seas around India. Indigenous 3-D scanners will be attached to the 10cm radars at Calcutta and Madras to enhance observational capability, while wind and temperature observations of the stratosphere will be measured by rockets fired from Thumba, Sriharikota and Balasore.

MONEX-79 is a sub-programme of the First Global Garp Experiment (FGGE) of the GARP (Global Atmospheric Research Programme). It will coincide with the launching of FGGE on 1 May and will end in August when the monsoon abates over India. Rs 250 million will be spent on the experiment.

From Dilip M. Salwi in New Delhi

How safe is the fast reactor?

Fast breeder reactors are an essential part of a nuclear future — but in the light of the incident at Three Mile Island would they be safe? **Professor F. R. Farmer**, Safety Adviser to the UK Atomic Energy Authority, discusses how to assess their dangers

FAST breeder reactor safety is currently an emotive topic, but it should be placed in context: absolute safety can never be achieved, and most industrial activities carry some risk of causing harm to people and the environment. One aim of design and operation must be to keep the risk sufficiently low, and there will be argument as to what this implies.

The increased recognition of risk has rendered obsolete the concept of maximum credible accident, and currently leads more to a discussion of residual uncertainties in accident and consequence modelling, rather than a prolonged description of the devices installed to minimise the chance of accidents and a demonstration of their reliability. In following this trend, it should be understood that some uncertainties will always remain, and it will only be possible to establish an increased degree of confidence that the system and its behaviour are well understood. This is important when considering the early versions of new systems—reactors or aircraft; it is not possible to establish the same degree of confidence in the first as may be developed by the fifth or the tenth, and to accelerate this learning process there should be extensive exchange of safety information and collaborative programmes on safety issues on an international scale. To date collaboration has been good, and it is to be hoped that commercial consideration will not prevent this exchange.

Since the fast reactor core is much more compact than thermal reactors, the heat production per unit length or volume is greater. The cooling system has to cope with this, and fast reactors use liquid sodium, a very effective heat transfer medium operating at near atmospheric pressure.

The control of power in a fast reactor is easier than most thermal reactors. Both depend on the small fraction of delayed neutrons, but the more compact core of the fast reactor simplifies the procedures for controlling power because the effectiveness of any control rod extends virtually throughout the reactor and is only slightly affected by the position of other rods. Elaborate systems of sector control, as in large thermal reactors, are thus unnecessary. It is easy to provide redundancy in the control and shut off sys-

tems, since in a large reactor with 30 rods, the insertion of any five would shut it down. The aim in current designs is to achieve sufficient diversity.

The removal of heat from the fuel in a fast reactor is in some ways easier and in others more difficult than in water or gas cooled reactors, as indicated by the following examples.

As sodium reacts violently with water—and additionally, hydrogen as a neutron moderator must be kept away from the core—all designs have intermediate sodium-to-sodium heat exchangers, and the final sodium-to-water exchangers are placed some distance from the reactor. On the other hand, the normal operating temperature of the sodium is several hundred degrees below its boiling point so that there is a wide margin to accommodate temporary disturbances of power or flow. In reactors of the tank design, the whole primary circuit is immersed in several thousand tons of sodium; as a result the fission product decay heat can be absorbed for several hours, sodium flow being by natural convective circulation. It is yet to be shown that the transition from the flow conditions under power to convective flow of a shut down reactor can take place safely—this is still being investigated.

There are several misunderstandings about the dangers of fast reactors; for example it has been said that a fast

reactor may behave like an atom bomb. The fuel of current fast reactors contains plutonium plus a higher percentage of ^{238}U , which is a neutron absorber, however, and this fuel cannot explode.

Another fear, that fast neutrons make the fast reactor more difficult to control, is likewise without foundation. The micro-second lifetime of neutrons in a fast reactor, as compared with milli-seconds in a thermal reactor, has no bearing on normal operation in which control operates through the delayed neutron fraction. Under most accident conditions the response is not dominated by the neutron lifetime, but by the doppler coefficient of the U/Pu fuel which cancels most changes in reactivity (resulting from some accident conditions) by a corresponding rapid change in the fuel temperature. The reactor becomes less reactive as the fuel temperature increases.

It has frequently been said that the particular feature of a fast reactor is that the core is not in its most reactive configuration. This is true, but it is an oversimplification in a comparison with thermal reactors. The light water reactors in their normal operation are not in their most reactive state although for different reasons. The light water reactors can be rendered prompt critical by changes in their coolant; the pressurised water reactors by a rapid temperature drop of the water; the BWR by collapse of the steam voidage in the reactor core. These conditions are now deemed to be reduced to an improbable event by design measures. The same degree of confidence has not yet been established for the fast reactor, perhaps because designs are viewed more critically. There are two types of fault conditions peculiar to a fast reactor which increase reactivity, one being the rapid expulsion of sodium from the mid-core regions, and the other a collapse or compression of the core into a slightly smaller volume.

As stated earlier, it is not, and never will be, possible to show that accidents cannot happen—it is only possible to show that the accidents are unlikely and their consequences reasonably understood and tolerable in relation to their estimated frequency.

How might this be done? It is likely that in newly developed systems such as the fast reactor, two lines will be followed in parallel. First, effective measures must be provided to prevent the escalation of minor faults into major accidents; hence all major accidents will have a very low probability of developing. Second, even if the preventative measures fail, the ways in which accidents develop and are limited and contained, are sufficiently well understood through an adequate (even if not perfect) appreciation of the

What Harrisburg means for the Rasmussen report

"I have heard recent statements that the accident at Harrisburg further devalues the Rasmussen report. I do not see any evidence for these statements; the sequence of events may or may not be consistent with the report. An accident may reveal new features and it is important to find out whether any new, unperceived sequence or new phenomena have occurred. However, the estimate of accident frequency is less important than the detailed assessment of each reactor and support systems. The chief merit of Professor Rasmussen's report lies in the detailed studies and in the control that is then exercised to ensure that his assumptions are met." Professor F. R. Farmer.

phenomena involved.

How far can these lines be followed at present? The normal classification of accidents into errors of control, heat removal, and so on, appear to be well covered with redundant and diverse protective equipment. The adequacy of these provisions may be subject to some discussion, mainly to improve confidence in the lower levels of accident frequency. It seems that two troublesome areas remain—a degree of uncertainty in the structural integrity of the internal reactor structures, when requiring confidence at low frequency levels, and some uncertainty in the various possible modes of failure propagation within fuel sub-assemblies and their subsequent development. Currently, many methods of early identification of onset of failure are possible, but the application of all sensors could lead to over-instrumentation to such an extent that the reactor could never operate effectively.

These uncertainties are likely to be resolved, although not solved, with sufficient confidence to enable the next stage to proceed. The greater stumbling block, ideologically, lies in the second parallel line of argument—that if preventative measures fail, the accident can still be contained. The difficulty in this approach is the continued escalation of the assumptions defining the accident. There is seldom an upper limit, and there is no maximum credible accident as conveniently postulated for the development of the light water reactors. Rather than question too closely the initiating assumptions, it may be more instructive to study two important phenomena, the mode of fuel behaviour and core dispersion in

a 'whole core accident', and the conditions of violent or passive interaction between hot fuel and sodium, known as FCI (Fuel Coolant Interactions).

The 'whole core accident' has been under study in USA, UK, France, Germany, Japan and USSR for up to 20 years. The research and theoretical programmes have been, and continue to be, extremely large and complex. The object has been to find out how fuel behaves when heated rapidly in the range 2,000 to over 4,000°C in timescales from a few milliseconds to a second or so.

Various core dispersive mechanisms exist. For a start the core must disperse if the fuel is vaporised: this mechanism can be violent and is studied to find the dynamic and static loadings in the core region and its surroundings. For some accidents, molten fuel may be swept out of the core by the coolant, and the loss of a small fraction of fuel would terminate the reactivity excursion. A third dispersive mechanism exists through internal gas pressures generated in the fuel by fission products. Theoretically, this could be very effective and could prevent fuel vaporisation, or reduce its extent. This mechanism has only been studied in the last few years, but experiments are continuing.

Fuel-coolant interactions have also been studied for well over 10 years, using various materials, including tin, aluminium, uranium oxide, water, sodium, and freon. Some theoretical correlations have been suggested in recent years, but there is not yet a generally agreed model, and more work is required particularly with UO_2 and sodium.

The question now arises of when there will be sufficient understanding of 'whole core accidents' and 'FCI', and what constitutes 'sufficient understanding'. Returning to the proposition of a two-pronged approach to safety, the first step is to show that accidents can be identified at an early stage and safely terminated—this is likely to be developed with considerable confidence. The other approach is to study accident conditions assuming protective measures fail. This route is open-ended, for at no time will it be shown that any accident developing in any way with no safeguards can be tolerated by any one design. However, these studies should lead to some optimisation in core materials, core geometry and surrounding containment, so as to tolerate a wide range of possible accidents even if these are of low probability.

This present stage in safety analysis is described by Hafele (Fusion and Fast Breeder Reactors, IAEA, A-2361 Laxenburg, Austria RR-7-8 November 1976 (Revised July 1977)) as the third stage in the safety debate. It is instructive to follow his argument. The first phase (1944–1959) gave much attention to the short neutron lifetime, the relatively small fraction of delayed neutrons, the speed of control devices, and a positive power coefficient, basically because the reactors were small and used metallic fuel. The second phase, up to 1970, considered larger reactors with mixed oxide fuel. The debate was then around the sign and size of the doppler coefficient, then around superheat, the sodium void coefficient and FCI.

The overriding feature of the third phase is a widespread and multilevel approach of proving the elements of the design concept that arose from the second phase. In this connection, one has to realise that the nature of what is considered a proof has evolved very substantially in the years since 1970. Engineering judgement is no longer considered sufficient. The probability approach to reactor safety has been introduced and related to the concept of residual risks, extremely small in probability but very large in the perceived consequences. The focus of the scientific and technical attention of the third phase, is therefore the depth and credibility of the experimental and theoretical proof of the design concept. In many countries there is now a growing interest in the attempt to quantify confidence and the degree of uncertainty. This is likely to continue. □

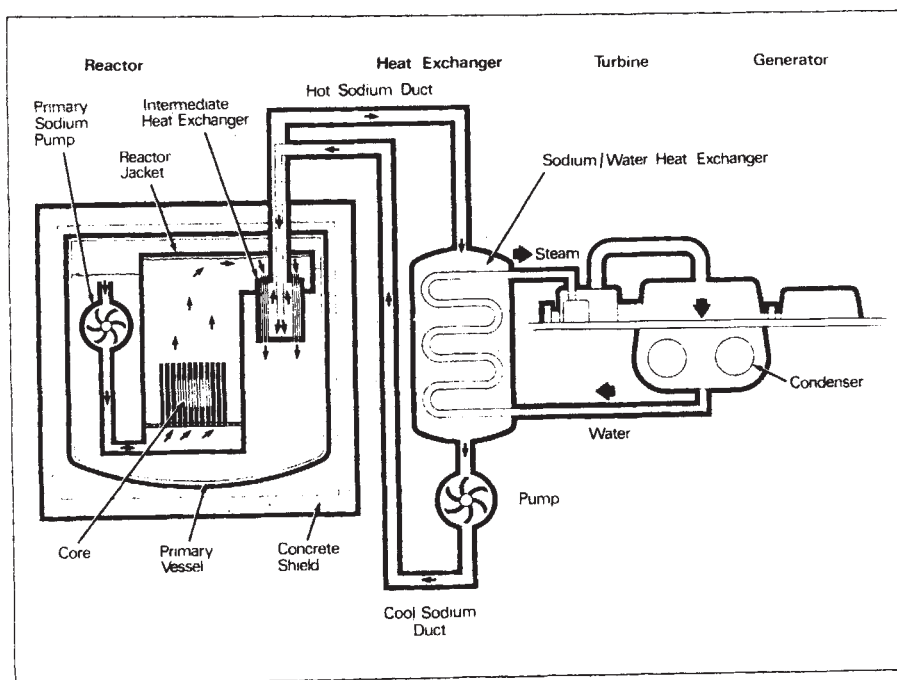


Diagram of the core and cooling system of the UKAEA's prototype fast reactor.

This article is based on a talk given by Professor Farmer at a conference on the politics of fast breeders at the end of last year. The conference proceedings are to be published by Macmillan.

How the West was scientifically explored

The US Geological Survey celebrates its centenary this year.

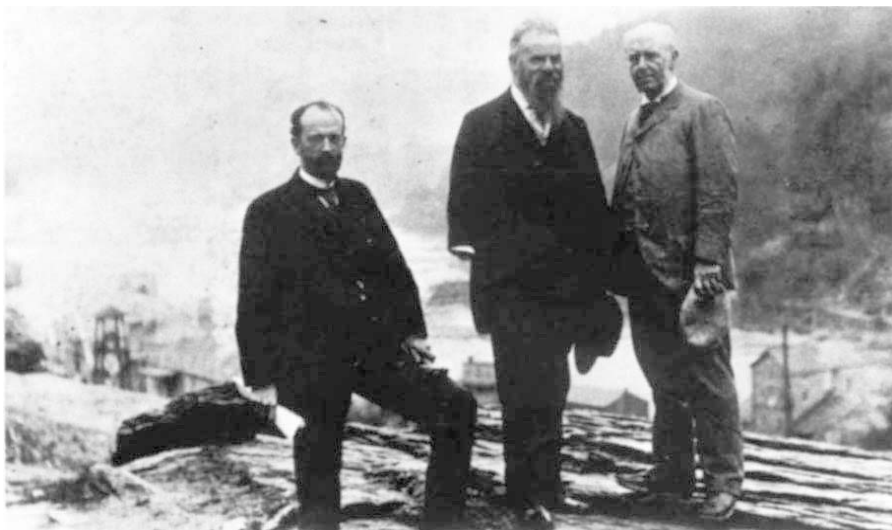
Harold L. Burstyn, the agency's official historian, surveys the past 100 years

At the close of the Civil War in 1865, only three and a half million people lived in the western half of the United States. A generation later, when the Superintendent of the Census of 1890 declared that he could no longer find on the map a continuous frontier line of settlement, more than eight and a half million Americans lived to the west of the first tier of states beyond the Mississippi River. In one of history's great migrations, Americans settled 430 m acres from 1870 to 1900.

With such a headlong rush westward of people seeking new opportunity on the frontier, exploration barely kept pace with settlement. From 1867 to the end of the 1870s, four separate groups, known collectively as the Territorial Surveys, mapped the west to uncover its mineral resources and assess the uses for its land. They were independent of the General Land Office, whose locally-hired surveyors divided the land into the rectangular plots acquired by settlers.

At a much higher level of precision, the Coast Survey, founded by Jefferson in 1807 as the federal government's first scientific agency, mapped both coasts by primary and secondary triangulation. The need to connect the Pacific to the Atlantic network led to a triangulation across the country and the change of the agency's name in 1878 to the Coast and Geodetic Survey. This change took place at a time of hope for political reformers anxious to restore the good name of the United States and the Republican party after the corruption of the Grant administration. President Rutherford B. Hayes and Secretary of the Interior Carl Schurz hoped to achieve their twin aims of reform and retrenchment by consolidating the Territorial Surveys into a single organisation and placing the parcelling of land in the hands of scientists.

On Joseph Henry's death in 1878 his place as leader of the National Academy of Sciences went to O. C. Marsh, the paleontologist. Under Marsh's direction, the academy prepared at the request of the forty-fifth Congress a plan for surveying and mapping the Territories. It recommended that the work be in a single agency, the Coast and Interior Survey, which would be the old Coast and Geodetic Survey re-



Survey director Charles Walcott (left) and John Wesley Powell, with Sir Archibald Geikie, their British counterpart, at Harper's Ferry, 1897.

named and transferred from the Treasury to the Interior Department. The geological work of the separate territorial surveys, three of them still in the field, would be combined into another agency, the Geological Survey. Finally, the academy report called for a federal commission, among whose members would be the heads of both surveys, to recommend changes in the laws concerning the public lands.

The final session of the forty-fifth Congress to which Marsh and his colleagues reported met after the election of 1878. At that time bills approved in the rush to adjourn each session often contained the only substantive legislation of the entire two years of the Congress, and appropriations bills (which had to pass if the federal government was to function) were the favourite vehicles for new departures. Thus the Sundry Appropriations Bill, the last to be passed by the forty-fifth Congress, attempted to reform the disposal of the public lands by placing their measurement and appraisal in the hands of scientists. But only two of the academy's proposals were in the Bill that President Hayes signed into law just before the Congress expired at midnight on 4 March 1879: the founding of the United States Geological Survey in the Department of the Interior for "the classification of the public lands, and examination of the Geological Structure, mineral resources, and products of the national domain"; and the formation of a commission to codify and propose revisions to the land laws. Both the Democratic-controlled House and the Republican-controlled Senate rejected the attempt to change the system by which the public lands were parcelled out to settlers. And when the commission on

the land laws made its recommendations two years later, they, too, were rejected by Congress.

To direct the new Geological Survey, Hayes and Schurz chose Clarence King, leader of the Geological Exploration of the Fortieth Parallel. From 1867 to 1873, King and his civilian colleagues had reconnoitered the natural resources along the route of the first transcontinental railway and had begun a series of monographs that set the highest scientific standards for the geology of the American west. King, "the best and brightest man of his generation . . . with talents immeasurably beyond any of his contemporaries", organised the new agency into districts, headquartered in Denver, Salt Lake City, and San Francisco. His topographers and geologists spread out to determine the character of the public lands, beginning with the major mining districts: Leadville in Colorado, Eureka and the Comstock Lode in Nevada, and the gold fields of California. King hoped that a decade of concentration on the principal ore bodies would lead to a rigorous theory of their formation. A sound geological understanding of the way metals were found in the earth might also show notorious wastefulness, while the work of the geologists could help investors in American mines sort out the favourable from the unfavourable prospects.

King wanted to extend the survey over the whole country, but the Interior Department was uncertain of the survey's mandate from the Congress to study those parts of the country that did not belong to the federal government, and it confined the survey to the public lands. Frustrated by petty politics and anxious to seek his fortune in mining, King resigned in 1881.



From the Survey's archives: earthquake damage in Charleston, 1886; geologist C. Willard Hayes on the limb of an exposed syncline in Maryland, 1897.

President Garfield replaced King with John Wesley Powell. Powell and his colleagues of the Geographical and Geological Survey of the Rocky Mountain Region had found the principles of a new approach to the earth, emphasising the role of erosion and deposition in forming the landscape. As an explorer he paid as much attention to the human as he did to the natural environment. He concluded that much of the interior lacked the rainfall to support traditional family farms. Dividing the public lands into uniform rectangles would invite monopoly by whoever controlled the banks of the streams. He called for a new system of land parcelling based on topography, for the setting aside of potential reservoir sites for public ownership, and for larger homestead grants in the dry areas. Powell's ideas lay behind the projected land reform of 1879 that the Congress had rejected.

Powell abolished the district offices, and based the survey's organisation on subjects such as glacial, volcanic, and other branches of geology, rather than territory. Determined to build a national agency, he used the Congressional authorisation, "To continue the preparation of a geologic map of the United States" to expand the survey to cover the whole country. Powell countered the "states' rights" argument against his federal agency by beginning in 1884 a cooperative programme of topographic mapping with Massachusetts, which, as one of the original states, had never contained any public lands. From Massachusetts, cooperation extended into other states and into geology and hydrology until it covered virtually the whole country and all its resources. After a broad investigation of economy in the federal government's scientific agencies, the Congress in 1886 renewed the survey's mandate for earth science.

Powell's chance to press for his programme of reform came with the droughts that began in 1886. By 1888,

after drought and blizzards had destroyed the open-range cattle industry and three consecutive dry seasons had brought despair to farmers in the central Great Plains, the Congress authorised Powell to conduct an irrigation survey. He immediately began an ambitious programme, including topographic mapping of a vast area, measuring the flow of its streams, and designating the sites for irrigation works and the lands they would irrigate. The Congress closed the remaining public lands to entry until Powell and his men were finished, but this provision generated such opposition to the irrigation survey that Congress abolished all of it except the designation of sites for reservoirs. At the same time the survey suffered the first of a series of budget cuts that finally drove Powell to resign in 1894.

Though the most drastic reductions owed more to both party politics and the Panic of 1893 than they did to opposition to Powell, his colleagues and successors in the Geological Survey never forgot the lesson of the 1890s. In the United States the primary purpose of federal civilian agencies for the physical sciences (until the National Science Foundation was created in 1950) has been to regulate. As a result, in times of constraint the regulatory function can usurp the resources devoted to research, however meagre they may be. From the failure of Powell's Irrigation Survey the Geological Survey's leaders learned that science could be saved only if administrative power was eschewed. For the next 30 years, they ensured that all possible regulatory roles were transferred to other agencies. As a consequence, its leaders believe, "the government's most productive research agency during the nineteenth century" has maintained its vitality in scientific research into the final quarter of the twentieth.

The trees of the United States passed rapidly into private hands until 1891

when the lame-duck fifty-first Congress empowered the President to create forest reserves by removing timbered lands from the market. Under Powell's successor, Charles D. Walcott, the Geological Survey mapped these national forests. Walcott also continued the Survey's inventory of the nation's water supplies.

As the vast fields of coal began to power the rapidly growing US industry, the Geological Survey began to test coals for their heating value and rocks for their use as structural materials. In 1908 the Congress asked the survey to look into mining safety, but since it was still averse to a regulatory role, these tasks were both transferred by the Congress to the Bureau of Standards and the newly-created Bureau of Mines.

Only in 1927, towards the end of the longest term of any of the survey's ten directors, George Otis Smith, did the survey abandon its opposition to becoming a regulatory agency. In 1925 Secretary of Commerce Herbert Hoover had wrested the Bureau of Mines out of the Interior Department into his own. The responsibility of the Bureau of Mines for supervising mining on public land was moved to the Survey's Land Classification Branch which became the Conservation Branch. With the growth of oil production on the outer continental shelf, the Geological Survey's Conservation Division is now its fastest-growing part.

The present Geological Survey continues to map the topography and the geology of the far-flung United States, to test its waters for quality and quantity, and to investigate the processes which formed the earth and the planets. The research ideal that has animated the US Geological Survey has borne fruit for a century. Yet the goal for which the Survey was founded—to make science the basis for the use of the land and its products—has not yet been reached. It remains the goal for the Survey's second century. □

correspondence

Dismantling DNA regulations

SIR,—I am delighted that John Maddox has taken up the debate on the problems related to regulations governing genetic manipulation (1 March, page 10). His analysis of the situation is perceptive when he states that the choice, if GMAG's procedures don't change soon, will be only whether to work with NIH guidelines here or elsewhere. In this regard John Maddox and I probably disagree only on time scale. He clearly feels that all may be resolved by the end of 1980 and asks if this is not enough, is it "a chance worth taking"?

If we are going to gamble we must know the odds and the stakes. At stake is the future of most genetic manipulation work in this country. The odds are heavily against anyone trying to continue in this highly competitive field if work has to be conducted under regulations so much more restrictive than elsewhere.

It is particularly bitter irony to find myself grouped with those scientists who invited constraints four years ago. That group was dominated by US scientists who are now working under relaxed NIH guidelines, while their colleagues here languish in the aftermath of their public outburst. From the outset, I have thought that those who invited constraints were doing both science and the public a huge disservice; I see no evidence to change my mind. A bureaucracy has been established and will be hard, if not impossible, to dismantle rationally.

In attacking the NIH guidelines, John Maddox subscribes to currently fashionable scare stories, those of "conjectured immunological hazards" and other problems "of expression". In fact when we talk of known proteins we can estimate possible dangers because we handle the proteins in their regular environment. We can therefore assess how to take precautions in handling the same products in genetic manipulation experiments. Such decisions, at least at the research laboratory level, are a matter of day-to-day care and not a matter for committee rulings and elaborate bureaucratic procedures. If there were risks which could be quantified and if there were a need for regulations which could be formulated rationally, then we might start to build from scratch. Unfortunately, we are not in a position to build, but rather in the position of dismantling excessively restrictive, empirically conceived, regulations.

If we were to accept the idea of dismantling the regulations by rational means we could fall into an expensive trap. Any conjectured risk would need to be quantified and the limit to conjecture and to expense involved in collecting the necessary data cannot be foreseen. What I asked for previously was an acceptance of the spirit of the NIH guidelines. Let us recognise the need for care when there is a known hazard, but let us not get lost in a morass of conjectured risks.

Yours faithfully,

ALAN R. WILLIAMSON

*Institute of Biochemistry,
University of Glasgow, UK.*

Mediterranean scientists stymied by bureaucrats

SIR,—Rather than focus on who attacked UN research programmes on the Mediterranean at the recent Geneva Conference of the Mediterranean States (15 February, page 506), and who succeeded in cutting three scientific research projects of major importance, one should analyse who failed (and why) to support the scientific component of the Mediterranean Action Plan (MAP). Incidentally, this component has been fully in line with the proclaimed "target areas" of the Group of 77 and of the UN system, among which the "reinforcement of scientific and technological capacities of the developing countries" is foremost.

That research results, produced by the 83-odd laboratories of the region, are still inferior in quality compared to those from (for example) Liverpool, Woods Hole, or Villefranche-sur-Mer, would hardly surprise anybody. Obviously, in recent years something different from competing with the centres of excellence has been attempted. Recently it became clear that this implicit aim has met with considerable success. At the November 1978 joint ICSEM-UNEP Review Meeting on the Mediterranean Pollution Research Pilot Projects, held at Antalya, Turkey, a large number of interesting contributions were reported. These came from many laboratories never before known to produce anything of international significance. UNEP's contribution to the execution of these programmes was in the form of equipment, training fellowships, and travel and subsistence support for such meetings. It never amounted to more than some 10% of the total on-the-spot expenditure for the participating laboratories' projects. However, its importance should be judged by the fact that it mobilised sizeable national resources for research in marine sciences in many developing countries; but, unfortunately, not in all. The latter, instead of asking for additional help in creating national research centres, have joined the ranks of critics of the programmes, although for different reasons.

The UNEP-sponsored and -catalysed multilateralism in scientific training, technical cooperation and data exchange, has indicated, for better or worse, the eclipse of the old northern paternalism. Even the well produced 'show-biz' contribution to the Antalya meeting by one of the most developed countries of the region, has not, and probably would not reverse this trend. It has become apparent that all around the Mediterranean there is emerging a broadly based international community of marine scientists, which is procuring the necessary atmosphere for appraisal of scientific data, for discussion of results, and for mutual understanding irrespective of present day political or economic disagreements. The bureaucrats, a large majority in the ranks of the 60 or so official governmental representatives at the Geneva Conference, have seen the old, more-or-less comfortable order vanish,

and see themselves increasingly made superfluous by the internationally oriented groups of scientists. The latter, as well known from experience, are difficult to control: they will soon be asking for more money, and will generally become a pain in the neck. Thus the hatchet was used, not surprisingly, on those who have produced the only tangible result in the four years of this unique exercise called MAP. Conversely, and in line with this thinking, a boost has been given to appealing, but vaguely defined, future plans. Among these, as Dr Walgate has correctly indicated, the development of alternative energy sources by the less developed countries has been accepted as one of the short term priorities.

The Geneva Conference was, in a way, a scaled down rehearsal for the UNCSTD. The message is clear and fully supports Moravcsik's admonition: do not let scientific and technological issues be decided by bureaucrats alone.

Yours faithfully,

VELIMIR PRAVDIC

*Ruder Boskovic Institute,
Zagreb, Yugoslavia.*

Cooperation in German health programme

SIR,—In his article about research in Germany (8 February, page 423) Robert Walgate refers to the federal government's programme on promoting research and development in the service of health 1978–1981 and to some worries of officials of the German research society (DFG). I am afraid there has been a misunderstanding, since there is no conflict between DFG and the federal ministries responsible for the health research programme.

When setting up the government programme, DFG and the federal government agreed that a close cooperation should take place. This is routine by now, DFG officials taking part in assessment meetings and quite often making valuable suggestions. In addition to this we have asked DFG to inform the government whenever they have the feeling that a project in health research is being funded by the federal government although the scientific quality does not seem to be sufficient. No example has been reported so far.

It is obvious that only research of good scientific quality should have an impact on government policy: poor research would do more harm than good. However, the federal ministries reserve the right not to fund a research project if it is likely that possible results would not have any bearing on the work of the ministries. In these cases, which happen quite often, the applicants are advised to submit their proposals to DFG or other institutions which do not have the same restrictions as we have, and whose financial means are not changed by the arrival of the new programme.

Yours faithfully,

HANS P. LORENZEN

*Federal Ministry for Research and
Technology, Bonn, West Germany.*

Plant protection

SIR,—Your correspondent writing on 'European plant health' (4 January, page 16), draws attention to and criticises certain aspects of the policy of the European and Mediterranean Plant Protection Organisation (EPPO) in listing organisms of plant quarantine significance for the EPPO region and describing them in its data sheets.

It is the function of EPPO to advise member governments on the technical, administrative and legislative measures necessary to prevent the introduction and spread of pests and diseases of plants and plant products. This it does through the agency of its working party on phytosanitary regulations, made up of specialists from nine of the 35 EPPO countries. The working party is, with the full approval of Council, developing procedures in three distinct areas. The first of these is the formulation of lists of non-European (A1) and European (A2) pests, based on extensive research data and the common experience of the plant health services of 35 member countries. The purpose of these lists is to draw the attention of countries to the organisms considered of greatest quarantine importance to the EPPO region so that appropriate precautions are taken in the form of stringent regulations. The second stage is the preparation, for each listed organism, of 'specific quarantine requirements', which EPPO countries are recommended to use as a basis for their phytosanitary regulations. These requirements are addressed specifically to the exporting country and specify conditions which have to be met for the importation of particular plants and plant products, based on the most detailed knowledge available on ease of detection of quarantine organisms in consignments and in the field, on persistence in soil, and so on. It is the formulation and acceptance of these recommendations as a basis for phytosanitary regulations which is the backbone of EPPO's international strategy to ensure that correct and adequate precautions are taken to prevent as far as possible the introduction of new pests without bringing the international exchange of commodities to a standstill. The third element is the preparation of data sheets on the A1 and A2 organisms. These sheets are addressed to a wider readership, including on the one hand the plant inspectors of importing and exporting countries and on the other, the general scientific public, which should be better informed on these organisms, many of which are poorly documented at present.

Further details on the activities of EPPO and its working parties will appear in the proceedings of a conference (Plant Health: the Scientific Basis for Administrative Control of Plant Parasites, ed. D. L. Ebbels & J. E. King) to be published by Blackwells for the Federation of British Plant Pathologists.

At all stages in the formulation of the A1 and A2 lists and in the preparation of specific quarantine requirements and data sheets, extensive consultations have been made with research specialists in the EPPO region and throughout the world. The exact status of some potentially dangerous diseases of virus, or virus-like, origin is less certain than that of most of the organisms on the EPPO lists. These uncertainties, which will doubtless be resolved by further research, are clearly stated in the EPPO data sheets and will similarly be reflected in the appropriate

specific quarantine requirements. The problem of new strains of quarantine organisms is a very difficult one, which has been thoroughly considered by the working party and can only be resolved at present by applying requirements to all those entities which are indistinguishable in inspection practice from the dangerous ones, or, in extreme cases, by prohibiting the importation of certain plants. It is impossible to foresee the appearance elsewhere in the world of a new aggressive strain of a pathogen already present in a country. Phytosanitary regulations can only be formulated in practice in terms of what is known to exist, not what might exist or appear. This also applies to pesticide-resistant strains, which may cause considerable trouble when introduced into new areas.

We believe that EPPO's phytosanitary strategy, which has been formulated by and has the support of the plant health services of its member countries in consultation with their own specialists, is a sound basis from which to tackle the practical problem of excluding dangerous pests from the region at a time when increasing quantities of exotic products from many parts of the world are being shipped to Europe in fast-moving freight transport. The A1 and A2 lists are under constant review and will be extended as necessary to include other quarantine pests as information on their importance becomes available. If the formulation of the lists and the publication of the data sheets encourages research workers to try to resolve some of the uncertainties about these pests, so much the better.

Yours faithfully,

A. H. STRICKLAND
G. MATHYS
P. JOURNET
JAC. FJELDDALEN

EPPO, Paris, France.

Safety in the news

SIR,—Perhaps some other readers noticed a remarkable coincidence of news items in the 8 March issue of *Nature*.

First there is the ban, after a ten-year struggle, of the 2,4,5-T herbicide in the US, because of the presence of dioxin. Then there is the recent release of the file on past nuclear reactor incidents, with the official comment "the absence of more serious effects is largely the result of luck"; and the occurrence of accidents with a Rasmussen probability of 10^{-12} per year. Also, there is the new official concern in the US with the potential dangers of low levels of radiation previously treated as safe.

Finally, J. D. Watson calls for the deregulation of all DNA research, not merely that on a laboratory scale of the past few years, but also the industrial-scale research of the future.

Yours faithfully,

J. R. RAVETZ

University of Leeds, UK.

Detecting frozen lunar gases

SIR,—The possibility of sending a lunar satellite into a polar orbit round the moon is being discussed. This will allow the observation of the interiors of craters near the poles, the bottoms of which may never have been reached by the sun since their time of formation. I have suggested earlier (*Phys. Bull.* 24, 312 and 453; 1973)

that in such cold crater bottoms one might find a continuous record of the evolution from the moon's interior of gases such as water, carbon dioxide, and many other materials volatile at lunar day-time temperatures.

In the absence of a significant permanent atmosphere on the moon, individual molecules will travel over a large part of the lunar surface; any which are not swept away by the solar wind or by becoming ionised will eventually find their way to a cold crater and freeze out there. The deposited strata of condensed gases cannot be studied without the landing of appropriate instruments, but the orbiter may make it possible to identify likely places for this.

Over the whole lunar surface, one could expect the top metre of regolith to contain something like fifty million curies of uranium. A large part of the radon-222 derived from the decay of this should be able to escape and with a $3\frac{1}{2}$ day half-life has a good chance of reaching the polar region before it decays. An atom of radon would be held in a sunless area and such an area will therefore concentrate the radon itself and its daughters from a much larger area. An observable amount of alpha-activity might therefore be concentrated although good spatial resolution would be needed to observe it. Detection would be made easier by the fact that the emission would be effectively from an infinitely thin source since the cold condensed layer would be thick enough to stop all alpha-particles from the underlying long-lived alpha-emitters. It might, however, be worth looking carefully at the energy distribution of the alpha-particles of polonium-210 which, together with its parent lead-210, could be expected to have remained in situ for some decades, during which there might have been some deposition of a detectable thickness of water if there had been any significant emissions of this during the period.

Over most of the lunar surface one would of course expect a thick-source spectrum derived from alpha-emitters still within the regolith with only a small thin-source contribution from the radon in the "gas phase" above the surface. A limit to the relative size of this may be estimated. The mean thermal velocity of ^{222}Rn at 0°C is about 160 m s^{-1} . Leaving the moon's surface normally with this velocity and without collision it would fall back to the surface again in 200 seconds. An average excursion would take about two minutes during which time it would travel horizontally about 20 km. In an average lifetime of five days it could thus make some 3,500 returns to the surface, and a cold crater bottom could expect to collect all the radon from 3,500 times its own area.

This figure must certainly be too high since it assumes an effective albedo of one, or zero resorption by the rest of the surface. This is quite unrealistic. The time of dwell on or in the surface after each return is not known but is clearly not zero. On the other hand, if the satellite could be made to travel well below the top of the radon atmosphere (half height nearly 10 km) the background of thin-source particles would be usefully reduced. All in all, it would certainly seem worthwhile to look for possible anomalies of alpha-emission near the poles.

Yours faithfully,

J. H. FREMLIN

Department of Physics,
University of Birmingham, UK.

news and views

Acetylcholine receptor clusters

from Eric Frank

THE acetylcholine receptor, a membrane protein that mediates transmission of information between nerve and muscle, has been extensively studied both physiologically and biochemically. Yet despite the wealth of available information concerning its distribution, pharmacological properties, synthesis, degradation and subunit structure, one still knows very little of how receptors come to be so precisely located at neuromuscular junctions. Small clusters of ACh receptors found on non-innervated muscle cells resemble the high density patches at synapses, and a knowledge of how these receptor clusters are formed may provide some insights into the development of the postsynaptic muscle membrane.

In mature muscles, the sensitivity to ACh falls 50 to 100 times when the point of ACh application is moved only 2 μm away from a nerve terminal, and within 100 μm the sensitivity is down by more than a factor of 1,000 (Kuffler & Yoshikami *J. Physiol. Lond.* **244**, 703; 1975). Anatomical studies using α -bungarotoxin as a specific, high affinity ligand for cholinergic receptors in vertebrate muscle have shown a similar degree of localisation of the ACh receptor to neuromuscular synapses (Barnard, Wieckowski & Chiu *Nature*, **234**, 207; 1971; Hartzell & Fambrough *J. gen. Physiol.* **60**, 248; 1972; Albuquerque, *et al.* *Proc. natn. Acad. Sci. U.S.A.* **71**, 1376; 1974; Fertuck & Salpeter, *J. Cell Biol.* **69**, 144; 1976). The distribution of ACh receptors on muscle fibres without a well established innervation is markedly different, however. Denervated, embryonic and neonatal muscles all have high densities of receptors over their entire surface (Katz & Miledi *J. Physiol. Lond.* **162**, 393; 1962; Diamond & Miledi **170**, 389; 1964). Until recently, extrasynaptic receptors in denervated and embryonic muscle

were thought to be uniformly distributed in the muscle membrane. With the idea of membrane as a fluid mosaic (Singer & Nicholson *Science*, **175**, 720; 1972), a reasonable model was that a motor nerve somehow kept the receptors organised in a cluster at synapses, but that in the absence of this neural influence, the receptors were randomly distributed throughout the membrane.

It came as some surprise, therefore, when ACh receptors on muscle cells cultured in the absence of nerve cells were found to be clustered (Vogel, Sytkowski & Nirenberg *Proc. natn. Acad. Sci. U.S.A.* **69**, 3180; 1972; Fischbach & Cohen *Devl Biol.* **31**, 147; 1973). The clusters were present even if special precautions were taken to ensure that the muscle cells had not been contacted by nerves *in ovo* before they were isolated and put into culture (Beckoff & Betz *Science*, **193**, 915; 1976). Since the discovery of receptor clusters on myotubes *in vitro*, they have been observed in extrasynaptic regions of denervated mammalian muscle *in vivo* as well (Ko *et al. Science*, **196**, 540; 1977).

Characteristics

Clusters of ACh receptors, often called hot spots because of their high sensitivity to ACh, are typically round or ovoid patches about 5-10 μm across. In culture, each myotube usually has at least one cluster, more commonly two or three, and two clusters can occur in close proximity. Patches occur both at the edges of cells and in the thicker, more central regions where they are frequently associated with a muscle nucleus (Fischbach & Cohen *Devl Biol.* **31**, 147; 1973; Anderson *et al. J. Physiol. Lond.* **268**, 731; 1977). They are found on both cell surfaces, either juxtaposed to the substrate or on the free surface. Most clusters have discrete boundaries; the receptor density falls to background levels within a couple of micrometres. Estimates of receptor density within a cluster made using radiolabelled α -bungarotoxin are

two-to-threefold lower than those at mature synapses (Sytkowski *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 270; 1974). Freeze-fractures of identified hot spots also show a particle density two or threefold less than at neuromuscular synapses (Yee *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3007; 1978). In myocytes of *Xenopus*, receptors in hot spots are grouped into many tiny spots or fine lines (Anderson *et al. op. cit.*). Given this uneven distribution, the peak density within a hot spot may be as high as in mature postsynaptic membrane. One ultrastructural study failed to demonstrate any unique membrane or cytoplasmic specialisation at receptor clusters, but the immunoperoxidase method used may have obscured some fine details (Vogel *J. Cell Biol.* **69**, 501; 1976).

Hot spots increase in number throughout the first week in culture (Sytkowski *et al. op. cit.*). Once formed, they are stable structures. Only rarely does a cluster move or disappear on a morphologically stable muscle cell (Anderson *et al. op. cit.*; Frank & Fischbach *Cell and Tissue Interactions*, (eds Lash & Burger) 1977), although their total number does decrease in older cultures (Prives *Cell*, **7**, 543; 1976). After the initial stage of cluster formation, the appearance of a nerve fibre has not been observed on morphologically stable muscle cells. This positional stability is correlated with a lateral stability of individual receptor molecules in a cluster. By labelling receptors with fluorescent α -bungarotoxin, Podleski and his collaborators showed that receptors within a cluster are not free to migrate within the plane of the membrane, whereas receptors not associated with a cluster can diffuse quite readily (Axelrod *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 4594; 1976). Clustered receptors are not metabolically stable, however. Twenty-four hours after receptors are irreversibly blocked with toxin, hot spots reappear at their previous locations (Frank *et al. Neurobiology*, (eds Otsuka & Hall)

Eric Frank is Assistant Professor of Neurobiology in the Department of Neurobiology, Harvard Medical School, Boston.

1979). Thus individual receptors within a cluster are being constantly replaced even though the cluster as a whole is stable.

Receptor clusters share several characteristics with the subneural patches at nerve-muscle synapses in culture. Both types of cluster are positionally stable over time, although they can disappear if the muscle cell is innervated at another location. Individual receptor molecules are present at a similar density and are metabolised relatively rapidly, with a half-time in the membrane much shorter than at adult neuromuscular junctions but comparable to that of non-clustered receptors (Schuetze *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 520; 1978). The only consistent difference that has been noted is that in cultures of nerve and muscle from *Xenopus* larvae, neural patches tend to be long and narrow, similar in shape to the overlying neurite, while aneural patches are circular or oval.

Significance of receptor clusters

Almost as soon as receptor clusters were discovered, it was hypothesised that they might serve as recognition sites for cholinergic nerves innervating muscle (Fischbach & Cohen *Dev Biol.* **31**, 147; 1973; Vogel *et al. Proc. natn. Acad. Sci. U.S.A.* **70**, 270; 1973). Under a variety of conditions, neuromuscular synapse formation is correlated with a high density of ACh receptors, both during embryonic development and during the reinnervation of denervated muscle. Clusters of ACh receptors on cultured muscle cells therefore might act as preformed postsynaptic sites. A cholinergic nerve, encountering a receptor cluster would form a synapse there.

This idea, however, turned out to be wrong. Work from two different laboratories using different animals and different methods are in full agreement that most, if not all, nerve-muscle synapses do not form at pre-existing receptor clusters. M. Cohen and his collaborators used fluorescently labelled α -bungarotoxin to follow the distribution of receptors during synapse formation on myocytes cultured from *Xenopus* larvae (Anderson *et al. op. cit.*; Anderson & Cohen *J. Physiol. Lond.* **268**, 757; 1977). The fluorescent patches under nerve processes, presumably at synapses, were very different in shape from those on myocytes uncontacted by nerve cells. In a few cases, the distribution of receptors was followed on a single myocyte while it was contacted by a neurite. Receptor molecules already labelled with fluorescent toxin coalesced under the neurite, and adjacent pre-existing fluorescent patches often disappeared. Neural clusters were thus formed by re-

cruitment of surrounding unclustered receptors, and perhaps by receptors from dispersed pre-existing clusters as well. Similar results were obtained in G. D. Fischbach's laboratory where the ACh sensitivity profile of chick myotubes in culture was determined physiologically during innervation by explants of embryonic chick spinal cord (Frank & Fischbach *Neurobiology*, (eds Oosuka & Hall) 1979) New areas of high sensitivity appeared at newly formed synapses. In no case was a synapse observed to form at a pre-existing sensitive spot. These pre-existing clusters often disappeared when they were adjacent to synapses, just as observed in the cultures of amphibian muscle. Because of the method used to assay ACh receptor distribution, it could not be determined whether receptors were recruited from adjacent areas or represented the incorporation of new receptors into the postsynaptic membrane.

It should be stressed that these findings do not bear on the importance of ACh receptors during synapse formation. There are appreciable numbers of receptors everywhere on these cultured muscles, at a density comparable to that in extrasynaptic regions of denervated, supersensitive muscle. Nor should one conclude that there are no preferred sites of innervation. If such sites do exist, however, they are not characterised by clusters of ACh receptors.

Another possibility is that receptor clusters on non-innervated muscles do not serve any useful function but are produced by the same mechanisms that operate during synapse formation to form high receptor densities in the postsynaptic membrane. During synapse formation, a nerve could simply trigger off an innate mechanism in the muscle resulting in the creation of a subneural cluster. Under normal conditions *in vivo*, motor nerves are present at very early stages of myogenesis, so that receptor clusters might normally appear only at synapses. But in culture, where muscle can be grown in the absence of nerve cells, cluster formation might be inappropriately triggered by other factors. Consistent with this idea is that synaptic and aneural receptor clusters in culture appear to be very similar to each other.

How nerves induce receptor clusters is unknown. An old theory was that ACh itself acted as a trophic agent, keeping the receptors clustered at the synapse. More recently, it has been suggested that the mere physical presence of the nerve could produce a high receptor density, since a thread or dead nerve placed on the surface of a muscle can produce a local area of high sensitivity. However, recent experiments by M. Cohen show that

clusters do not form under sensory nerves (personal communication). The most promising current idea is that motor nerves release a chemical substance other than ACh that induces receptor accumulation and/or cluster formation. Three laboratories have reported that extracts of nervous tissue added to muscle cultures increase the number of ACh receptors (Podleski *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 2035; 1978; Christian *et al.* **75**, 4011; 1978; Jessel *et al. Neuro Abs.* **4**, 369; 1978) and at least one of these extracts increases the number of receptor clusters as well. The isolation of the relevant substance in these extracts may provide an important tool in the study of the molecular mechanisms of receptor clustering. □

Cell lineage in early development

from Jonathan Slack

THE days of descriptive biology are often thought to be over, but in the field of experimental embryology there is a type of pure observation which retains its importance and which continues to pose considerable technical difficulties. This is the construction of the 'fate map' for an animal embryo, which shows what will happen to each region of the embryo during normal development. An accurate fate map is essential for the correct interpretation of experiments which seek to establish the causal relations of development by interfering with the normal processes.

Only a limited amount of information can be obtained by the direct observation of embryos because of the complexity of cell numbers and morphogenetic movements, although it has been possible to trace the cell lineage of the nematode *Caenorhabditis elegans* up to 182 cells by Nomarski optics (Deppe *et al. Proc. Natn. Acad. Sci. U.S.A.* **75** 376; 1978). So the technical problem is that of specifically labelling a certain region of the embryo in such a way that the progeny of the marked cells can subsequently be found and identified, and to do this without damaging the marked cells or otherwise perturbing normal development. The methods used in the past have included staining with vital dyes, application of extracellular carbon particles and insertion of homologous

Jonathan Slack is a developmental biologist working at the Imperial Cancer Research Fund's Mill Hill Laboratory, London.

grafts whose cell nuclei are labelled with ^3H -thymidine or with a chromosomal marker. Much has been done with these methods, particularly on the amphibian and chick embryos, but each has its own limitations and disadvantages and in particular it is not often possible to mark single cells in the early embryo.

A technique has recently been introduced which seems to offer considerable scope for the construction of accurate fate maps for several types of embryo and which is particularly suitable for the marking of single cells. It is described in a paper by D. A. Weisblat, R. T. Sawyer and G. S. Stent (*Science* **202**, 1295; 1978). The marking is carried out by the injection of horseradish peroxidase (HRP) into one of the large cells arising from an early cleavage division (blastomeres). The embryo is then allowed to develop for a certain time during which the enzyme persists passively in the cytoplasm of the daughters of the injected blastomere. When the embryo has reached the desired stage it is fixed and the regions containing the peroxidase can be visualised by standard histochemical procedures, either in whole embryos, if they are small enough, or on frozen sections. In this paper the fates of various early blastomeres of the leech *Helobdella triserialis* are traced. The enzyme persists in the appropriate clones for several days in an active form and it does not seem to interfere with normal development. HRP has a molecular weight of 40,000 and would not therefore be expected to pass from cell to cell through gap junctions. By contrast it is shown that a low molecular weight fluorescent dye, lucifer yellow, does spread to the whole embryo after injection into a single blastomere.

The technique has been applied to the early amphibian embryo by M. Jacobson and G. Hirose (*Science* **202**, 637; 1978; *Zoon* **6**, 149; 1978). When one of the first two blastomeres of *Xenopus laevis* is injected the whole of the appropriate side of the embryo can subsequently be shown to be labelled. In only one organ is there any contribution of cells from the contralateral side and this is the ventral part of the retina which is found to consist on both sides of a mixture of labelled and unlabelled cells. By means of injections of single blastomeres at 4, 8 and 16-cell stages it is shown that only the dorsal blastomeres contribute to the neural plate and that within the neural plate the clones make up longitudinally arranged strips.

These authors also make use of mosaic embryos which are produced by subjecting newly fertilised eggs to a heat shock. This causes suppression of

the second polar body and in a small proportion of cases leads to a situation in which the first two blastomeres contain different numbers of chromosome sets. This difference of ploidy is visible in the progeny clones as a difference of cell size and nuclear staining intensity. The results obtained from this method with respect to the shapes of clones and the provenance of the retina are substantially the same as those obtained from the HRP procedure, but the introduction of mosaics for the fate mapping of amphibia focuses attention on a rather important aspect of the interpretation of these maps. This is highlighted by another technique known as gynandromorph mapping which has been used to produce fate maps for the blastoderm stage of *Drosophila simulans* and other insects (reviewed by Janning in *Genetic Mosaics and Cell Differentiation*, Springer-Verlag, 1978). This also depends on the production of a mosaic after the first nuclear division, but in contrast to *Xenopus* in which the first cleavage is usually in the median plane for all embryos, in the insect the first

nuclear division is randomly oriented. So the early stages have no definable fate map because different individual embryos are different, and the mapping is carried out with respect to the later, blastoderm, stage by statistical inference from the frequencies with which adult structures appear within the same clone.

This dramatic difference between two types of animal embryo: one with reproducibly oriented early divisions and the other without, should help clear up a misconception often found in the literature. It is often assumed that an account of a cell lineage which does not vary between individual embryos gives some information about the decisions made during development and the states of determination of the cells. But it does not. As fate mapping is a purely descriptive procedure, additional investigations must always be made to establish whether clonal boundaries separate regions of different intrinsic cellular characteristics or whether they are simply passive consequences of the orientation of early cleavages. □

Specificity of DNA uptake in bacterial transformation

from J. R. Saunders

THE genetic transformation of bacteria by purified DNA has been of great significance in the development of molecular biology and is an essential technique in recombinant DNA research. However, little is known of the overall importance of transformation as a means of genetic exchange between wild populations of bacteria. Some bacterial species such as *Streptococcus pneumoniae*, *Bacillus subtilis*, *Haemophilus influenzae* and *Neisseria gonorrhoeae* become naturally competent (able to bind and take up DNA into a DNase-resistant form) in particular growth conditions. This suggests that gene transfer by transformation may be a relatively frequent event in the biology of these organisms.

B. subtilis or *S. pneumoniae* can take up DNA from unrelated (hetero-specific) organisms as well as homo-specific DNA. On the other hand *H. influenzae* can discriminate against foreign DNA and will only take up *Haemophilus* DNA efficiently (Scocca *et al.* *J. Bact.* **118**, 369; 1974). Two elegant papers from Hamilton Smith's laboratory at Johns Hopkins have re-

cently provided some exciting insights into the mechanisms involved in the specificity of DNA transport during *Haemophilus* transformation. Sisco and Smith (*Proc. natn. Acad. Sci. U.S.A.* **76**, 972; 1979) have examined transformation specificity in *H. influenzae* using as a probe an 8.1 kilobase fragment of *H. parainfluenzae* DNA cloned in *Escherichia coli* on the plasmid pBR322. This fragment was adsorbed very well by competent *H. influenzae* cells whereas pBR322 DNA was not. Recognition of *Haemophilus* DNA could not possibly involve conventional restriction/modification systems because the probe DNA had been propagated in *E. coli* and would have lost the specific *Haemophilus* modification pattern. It seemed likely therefore that *Haemophilus* DNA contains a special base sequence which determines uptake. Sisco and Smith then located these putative sites of specificity on two smaller fragments (120 and 140 base pairs long) lying 3.8 kb apart on the restriction endonuclease cleavage map of the 8.1 kb fragment. Competition for binding to *H. influenzae* cells between labelled 8.1 kb fragment DNA and chromosomal DNA from *H. influenzae* or *H. parainfluenzae* indicated that the *Haemophilus* genome contains

J. R. Saunders is a Lecturer in the Department of Microbiology, University of Liverpool.

about 600 copies of this uptake sequence (approximately 1 site per 4.0 kb). This specificity site does occur in certain heterologous DNA species but only at a low frequency (about 1 site per 300 kb). This limits successful uptake to fragments of foreign DNA carrying the recognition sequence. The actual recognition site is estimated to be only 8–12 bases long but the exact sequence awaits elucidation.

The ability to recognise the uptake sequence apparently resides in a membrane protein which is capable of specific binding of *Haemophilus* DNA and which is found only in competent cells (Deich & Smith in *Transformation* 1978 (eds. Glover & Butler) 343, Cotswold Press, Oxford; 1979). This protein could merely be a specific surface receptor for DNA but it might also act as a pilot protein guiding homologous DNA into the cells. It is worth noting that a surface-located endonuclease is required for transformation of *S. pneumoniae* where one strand of double-stranded transforming DNA is degraded whilst the complementary strand is drawn into the competent cell (Lacks in *Microbial Interactions. Receptors and Recognition. Series B.* 3, 179; 1977). However, the situation must be different in *Haemophilus* because DNA remains double-stranded for several minutes after uptake and there is no evidence as yet that the receptor protein has nuclease activity.

A paper from the laboratory of Alexander Tomasz at the Rockefeller University, New York shows that *N. gonorrhoeae* also has a system for discriminating against foreign DNA during transformation (Dougherty *et al.* *Biochem. biophys. Res. Commun.* 86, 97; 1979). Gonococci will adsorb large amounts of both homologous (*Neisseria*) and heterologous DNA but only the former is taken up into a DNase-resistant (presumably intracellular) form. Competition experiments suggest that gonococci have two kinds of receptor on the cell surface. One receptor binds all types of DNA and the other is specific for DNA from closely related species. DNA is transported into the cell solely from the latter receptor. Whether or not the molecular basis for the specificity observed in *N. gonorrhoeae* is similar to that found in *Haemophilus* remains to be seen.

The specificity of DNA uptake exhibited by *H. influenzae* and *N. gonorrhoeae* raises some interesting possibilities as to the role of transformation in natural populations of these bacteria. Both organisms are prone to autolysis, releasing DNA, and both are highly competent, suggesting that transformation may often occur in the wild. Efficient transformation, however, poses its own problems since indiscriminate uptake of heterologous DNA on a large

scale could swamp restriction systems and compromise the genetic integrity of the host. Surface-located systems for distinguishing homologous DNA may therefore have evolved as additional genetic isolating mechanisms. Even so other bacteria which are also readily transformed *in vivo* such as pneumococci have maintained their species identity without the benefit of a discriminatory uptake system for DNA. The origin and evolution of specificity sites in *Haemophilus* together with possible similarities with restriction endonuclease recognition sites should become more apparent when their full base sequence is known. The specificity of *Haemophilus* transformation is, however, evidently yet another example of the ability of proteins to recognise defined sequences in DNA. □

Myofibrillar connection

from Pauline Bennett

It has long been suspected that in the striated muscle fibre there is a cytoskeleton which holds the constituent myofibrils in place. The presence of structural elements parallel to the fibre axis is indicated by the elastic resistance of the muscle to stretch which is not attributable to the contractile apparatus itself. Further, when the bulk of myosin and actin are extracted from myofibrils there remain 'ghosts' whose integrity is not lost even though in the light microscope only the Z-disks can be seen. The presence of transverse elements is suggested by the tendency for the Z-disks of neighbouring myofibrils to be in axial register across the fibre giving rise to the cross-striated appearance of the muscle cell. The components of this framework have not until recently been identified.

There have been suggestions, usually based on electron micrographs of muscle from which the contractile proteins have been extracted that a fibrous element is involved. For example, from such observations and an investigation of the components of extracted muscle, Maruyama and his colleagues (*J. Biochem.* 82, 317; 1977) consider that there is present a framework of longitudinal elastic filaments around each myofibril and that these filaments are composed of a protein, connectin. Connectin is however insoluble and has proved difficult to characterise.

In micrographs of intact muscle such

filaments are not in general observed, only membranous material is seen lying between the myofibrils. These membranes are obvious candidates for a supportive role. The sarcoplasmic reticulum sheaths each myofibril and could contribute to the longitudinal element whereas the transverse (T) tubule system is invaginated across the muscle fibre and could form the transverse elements. Furthermore the T-tubules occur at very precise positions with respect to the sarcomere structure, this position depending on the source of the muscle. In avian and amphibian skeletal muscle the T-tubules circle the myofibril at every Z-disk level and this relationship is maintained during contraction of the muscle. The question therefore is, what structural links are there, if any, between the membrane system and the contractile apparatus? One component of the transverse links seems now to have been identified.

Lazarides and his colleagues in a recent series of papers have been investigating the role and distribution of a protein they call desmin (from the Greek *desmos*, a link). This protein of chain weight 55,000 they first found in chicken smooth muscle. They have shown by indirect immunofluorescence that its antibody labels structures in all avian muscle cell types including the Z-lines of skeletal muscle fibres. Now Granger and Lazarides (*Cell* 15, 1253; 1978) have studied in more detail the three-dimensional distribution of antibody binding in striated muscle. To do this they have been aided by the valuable discovery of a method for obtaining sheets of Z-disks. Whole muscle, when homogenised, tends to be sheared lengthwise to form myofibrils. However, if most of the myosin and actin are first extracted from the muscle with KCl and KI the axial strength is reduced and blending shears the fibre across its width giving honeycomb-like sheets of Z-disks. SDS gel electrophoresis shows that these sheets are enriched in desmin compared with the original muscle although they still contain substantial amounts of actin, some α -actinin and traces of myosin and tropomyosin. Indirect fluorescent antibody labelling of ghost myofibrils shows that desmin, α -actinin and actin are all present at the Z-disk level. Similar labelling of Z disk sheets reveals that α -actinin is located as expected in the Z-disks whereas desmin labels the areas between the Z-disks forming an interconnecting network across the fibre. Antibody to actin seems to label the whole Z-disk sheet.

Previous observations of Lazarides and Granger (*Proc. natn. Acad. Sci. U.S.A.* 75, 3683; 1978) using a fluorescent dye to mark the hydrophobic,

Pauline Bennett is in the MRC Cell Biophysics Unit, King's College, London.

and hence presumably the membrane, positions in glycerol-extracted fibres showed substantial fluorescence at the Z-line level and around the Z-disks in Z-disk sheets. The location of desmin thus seems to coincide with that of the membrane, probably the T-tubule system.

Since the resolution of the light microscope does not allow the precise position of desmin to be determined, Lazarides and his colleagues suggest a number of possible interpretations of the data. First the desmin may be involved in mediating the adhesion of the T-tubule to neighbouring Z-disks. Alternatively, desmin might be associated with direct fibrous links between Z-disks which are independent of the T-tubules. The first possibility offers a good explanation for the precise location of the T-tubules in the muscle cell and finds support in the general observation of the authors that desmin is present where membranes and actin appear to be associated such as in the intercalating disk in heart muscle. A test of this hypothesis would be to study the distribution of desmin in mammalian skeletal muscle. Here the T-tubules are located at the junction of the A and I-bands, not at the Z-disk level as in avian muscle, which should lead to a different antibody staining pattern. Unfortunately Lazarides and Balzer (*Cell* 14, 429; 1978) have shown that their antibody is specific only for avian muscle, but they have some evidence from two-dimensional gel electrophoresis that a similar protein is present in mammalian muscle.

The alternative proposal that desmin forms direct links between Z-disks is based in addition on the observations of filaments lying between Z-disks in ghost fibres. However, these are considerably longer than the distance between myofibrils in intact muscle. The evidence that these filaments are composed of desmin is indirect. Lazarides and his colleagues consider that desmin is the same as the protein, skeletonin, isolated by Small and Sobieszek (*J. Cell Sci.* 23, 243; 1977) from chicken smooth muscle where it is the major component of the 100 Å filaments. There is a possibility that these two proteins are not identical. The extraction procedures used by the two groups are different, as is, slightly, the molecular weight attributed to the proteins. Furthermore it is unusual to find a protein which does not have a similar morphology in very different cell types and no 100 Å filaments have been detected in skeletal muscle, although a few could easily go unnoticed. If these objections are not valid can I broadcast a plea for the groups involved to get together to decide on a common nomenclature. In

view of the, at present, confused state of the field of 100 Å filaments, this is an important point to clarify.

Whatever the state of desmin in the muscle it is clear that in this protein we now have a handle with which to tackle the problem of how the transverse order is maintained in the striated muscle cell and can further hope to discover how it is related to the putative longitudinal cytoskeleton.

Enzyme therapy in genetic diseases

from Gregory Gregoriadis and Michael F. Dean

THERE are numerous genetic disorders in which, because of a partial or total absence of a specific enzyme, substrates normally metabolised by the enzyme accumulate intracellularly or extracellularly. This leads to profound defects in the function of the patient's tissues and in many instances to severe incapacitation or early death. Treatment for most of these disorders has been mainly symptomatic and, in a few cases, dietary. However, as knowledge of the molecular mechanism of these diseases progresses, more sophisticated attempts at therapy can be undertaken. Since therapy by genetic manipulations is only a distant possibility, efforts have concentrated on more immediate solutions, namely reduction of the substrate load by restoration of enzyme activity or other methods.

Enzyme therapy has recently gained a new impetus as a result of the advances in methodology discussed by D. Robinson (Queen Elizabeth College, London) which have provided us with bulk quantities of enzymes from human tissues, especially placenta (for example α -galactosidase, C. C. Sweeley, Michigan State University, East Lansing; glucocerebrosidase, G. L. Dale, City of Hope Medical Center, Duarte) and in the possible use of bacteria for the synthesis of biologically active proteins (L. Villa-Komaroff, University of Massachusetts Medical School, Worcester; R. D. Schmickel, University of Michigan, Ann Arbor). Another reason is the discovery that many enzymes possess recognition markers which facilitate their very rapid specific uptake by cells (E. Neufeld, National Institutes of Health, Bethesda). There is, for instance, a

phosphomannosyl recognition marker on several lysosomal hydrolases mediating their uptake by fibroblasts (W. Sly, Washington University School of Medicine, St. Louis), a fucosyl marker on glycoproteins recognised by hepatocytes (R. L. Hill, Duke University Medical Center, Durham) and an as yet uncharacterised group on β -hexosaminidase A which seems to promote binding of the enzyme by brain synaptic plasma membranes (J. W. Kusiak, National Institutes of Health). In the case of human β -hexosaminidase of which hepatic uptake in cats is due to a receptor recognising N-acetylglucosaminyl and mannosyl residues on the enzyme, unwanted participation of the liver can be suppressed (in favour of other tissues) by substances which compete for similar binding sites. This allows the enzyme to circulate longer and by increasing blood brain barrier permeability after exposing animals to hyperbaric oxygen, enzyme can thus reach nervous tissue (M. C. Rattazzi, State University of New York at Buffalo). Accessibility of enzymes to brain is also achieved by hypertonic shock using arabinose or mannitol which shrink the cells lining the vascular endothelium (J. A. Barranger, National Institutes of Health).

Difficulties in controlling the fate and effect of enzymes *in vivo* could be overcome by the use of carriers (naturally occurring or synthetic) which may, in various ways, improve enzyme distribution and action. Disappointingly, in type I Gaucher's disease in which glucocerebrosidase storage results in hepatomegaly (and splenomegaly), multiple injections of human glucocerebrosidase: β -glucosidase entrapped in γ -globulin-coated erythrocytes (E. Beutler, Scripps Clinic and Research Foundation, La Jolla) or in liposomes (G. Gregoriadis, Clinical Research Centre, Harrow) failed to reduce appreciably the size of the liver even though this tissue is the principal site for localisation of the carrier-associated enzyme. Nonetheless, after 3 years of treatment with liposomal enzyme there was some clinical improvement of the patient. Fortunately, neither of the carriers affected blood clotting factors or α_2 -macroglobulin (liposomes). Some clinical improvement was also seen in Gaucher's disease patients injected with the free enzyme but a substantial reduction of hepatic glucocerebrosidase (also shown in liver biopsies exposed to the enzyme *in vitro*) was not paralleled by a reduction in the organ's size (R. O. Brady, National Institutes of Health). It may be that long-term therapy is required for the reversal of the process of organ enlargement and, at any rate, the amount of enzyme given should be sufficiently high to at least prevent further glycolipid storage

The 2nd International Symposium on Enzyme Therapy of Genetic Diseases was held at Hilton Head Island, South Carolina on 4-7 March and was organised by R. J. Desnick under the sponsorship of the National Foundation-March of Dimes and the Mount Sinai School of Medicine.

Enzyme therapy by transplanting hepatic or bone marrow cells in immunosuppressed animals with model diseases, was partly successful (D. Sutherland, University of Minnesota, Minneapolis) although, with the exception of acute hepatic failure (where only temporary action is required), rejection remains a formidable obstacle for use in man. There has been, however, some success with bone marrow transplants in immune deficiency disease. Furthermore, transplants of HLA-identical fibroblasts increased levels of the relevant enzymes in patients with Hurler's, Hunter's and Sanfilippo type A diseases for up to several years and there was no evidence of graft rejection (M. F. Dean, Kennedy Institute of Rheumatology, London; D. Gibbs, Clinical Research Centre, Harrow). A more hopeful alternative would, of course, be transplantation of cells after they have been corrected for the malfunctioning or missing gene by one of the available methods.

A subtle and perhaps more efficient method of restoring enzyme activity in a large group of disorders is the administration of massive doses of specific vitamins (see review by C. R. Scriver, in *Treatment of Inborn Errors of Metabolism* (eds Seakins, Saunders & Toothill) 127, Churchill Livingstone, Edinburgh and London, 1973) which through a variety of mechanisms (for example the coenzyme may form part of the active site of the enzyme, affect apoenzyme synthesis and degradation or influence the specific activity of the enzyme by altering its conformation), result in beneficial biochemical and clinical changes (G. E. Gaull, Mount Sinai School of Medicine, New York). Much of the understanding of vitamin action has come from work with vitamin B and aspartate- β -decarboxylase (A. Meister, Cornell University Medical College, New York). A similar (but limited to mannosidosis) approach may be ingestion of Zn which can activate the partially missing α -mannosidase (G. A. Grabowski, University of Minnesota, Minneapolis). The almost certain effects of Zn on the absorption and metabolism of other trace metals could, however, prove unacceptable. In yet another approach advantage is taken of the equilibrium thought to exist between the various pools of substrate in the body. For instance, plasmapheresis coupled with a phytanic acid-poor diet lowered serum phytanic acid considerably and slightly improved muscular strength in a patient with Refsum's disease. Plasmapheresis alone reduced tri-

hexosylceramide levels to normal for a several day period in a patient with Fabry's disease (H. W. Moser, The John F. Kennedy Institute, Baltimore). The latter finding was also observed after repeated injections of plasma α -galactosidase A isozyme in two brothers with the same disease (R. J. Desnick, Mount Sinai School of Medicine).

Although some of these attempts at therapy have already benefited patients and some may have saved lives, most of the clinical trials were not preceded by animal experimentation, because animal models of human genetic diseases are extremely rare. The setting-up of such models is essential if therapy is to become more effective. An interesting example of a mouse with multiple lysosomal enzyme

deficiencies was demonstrated by P. Pentchev (National Institutes of Health). Even so, direct clinical trials are often a necessity. For instance, work with animal models is of little use in determining the likely immunogenicity of enzymes or cells prepared from human sources or whether cells from seemingly healthy donors may be contaminated with viruses, infectious viral components or with malignant cells. Some of these contaminants are impossible to measure or even anticipate and recipients, especially if immunosuppressed, may therefore be at risk. Similarly, artificial carriers such as liposomes, although controllable in terms of infectious contaminants, may exaggerate antigenicity through an adjuvant action



A hundred years ago

IN the interests of British science we have refrained now for some time from referring to the evil days which have fallen upon one of the most reputable of our learned societies. The time, however, has now come when silence is impossible. At the meeting of the Royal Astronomical Council yesterday, the Astronomer-Royal, in consequence of the recent action of the Council—an action inevitable when the present constitution of that body is considered—resigned his seat at the board. We cannot too much regret that this Society, the traditions of which are second to none in Europe, should have been utilised for some years past by an advertising clique who have everything to gain by their connection with a body of honourable students of science. The withdrawal of men long known for their astronomical work from the Council commenced some time since. It has now culminated in the resignation of the Astronomer-Royal, and we are informed that other resignations are to follow; indeed, a man of scientific repute risks somewhat in being found amongst the Councillors. Surely the Fellows of the Royal Astronomical Society of London are strong enough to remedy such a state of things as this.

At a recent meeting of the Medical Society of Berlin Prof. Virchow gave (by previous request) his views on the subject of the Plague of Astrakan. . . .

At the beginning of most pestilential epidemics a Committee of doctors has generally declared that it was not the plague. They pronounced it petechial typhus. This was the case immediately before the outbreak of the plague at Rescht, when the disease had been long confined in Kurdistan and Mesopotamia. M. Tholozan was the first to say it was plague, and that the case was that, not of a great epidemic, but of a latent disease, spreading slowly and attacking only a few. It is indubitable that we have there a true centre,

whence the disease gradually spread, and I do not see why we should go to India, where the disease has not prevailed for many years past. Proceeding logically, we shall accept this course: From Kurdistan and Mesopotamia to Persia and on to beyond the Caspian. Even if the present cases on the other side of the Caspian were accompanied by pneumorrhagia, I would not hesitate to say they belonged to the plague proper, and that the disease is the same as that in Mesopotamia. The symptoms are very different from those of petechial typhus, the disease which the Turkish doctors affirmed. If near Salonichi (Xanthi) there be really petechial typhus accompanied by *Metastasis bubonica*, I fear it is the plague. It remains to keep our eyes open and see what happens after the return of the Russian army from the infected country (an occurrence which may well rouse grave apprehensions).

What has been done for our protection is little apt to tranquillise us. A blockade comprising all the frontier as well as the coast, from the Baltic to the Black Sea, seems to me illusory.

One example of severe quarantine has occurred in this century in the case of the plague at Noja, in Bari (Kingdom of Naples), in 1815. Trenches were dug, and three cordons of sentinels were formed (the third round the entire province), with orders to kill whoever tried to break the blockade and did not stop at the first summons; and some individuals were actually killed. But I cannot think an entire country is able to protect itself thus. Examination of passports would be excellent if those who deliver passports and certificates of health were angels. But the Russian functionaries are men, and think like men. The impossibility of always getting true certificates of origin has been seen in the case of the cattle plague. I consider, however, that pressure should be exerted on Russia to form a blockade of the infected districts. And especially it should be seen to, that the returning Russian army does not bring any pestilential germs with it. As to restrictions on communications by land, the greatest of these are ineffectual for the end desired.

From *Nature* 19, 10 April, 543, 545; 1879.

G. Gregoriadis is at the Clinical Research Centre, Harrow and M. F. Dean at the Kennedy Institute of Rheumatology, London.

review article

The anthropic principle and the structure of the physical world

B. J. Carr* & M. J. Rees

Institute of Astronomy, Madingley Road, Cambridge, UK

The basic features of galaxies, stars, planets and the everyday world are essentially determined by a few microphysical constants and by the effects of gravitation. Many interrelations between different scales that at first sight seem surprising are straightforward consequences of simple physical arguments. But several aspects of our Universe—some of which seem to be prerequisites for the evolution of any form of life—depend rather delicately on apparent ‘coincidences’ among the physical constants.

THE structure of the physical world is manifested on many different scales, ranging from the Universe on the largest scale, down through galaxies, stars and planets, to living creatures, cells and atoms. Only objects such as quarks and leptons may be devoid of further substructure. Each level of structure requires for its description and explanation a different branch of physical theory, so it is not always appreciated how intimately they are related. We will show here that most natural scales are determined (to an order of magnitude) by just a few physical constants. In particular, the mass scale and length scale (in units of the proton mass m_p and the Bohr radius a_0) of all structures down to the atom can be expressed in terms of the electromagnetic fine structure constant, $\alpha = e^2/\hbar c$, the gravitational fine structure constant, $\alpha_G = Gm_p^2/\hbar c$ and the electron-to-proton mass ratio, m_e/m_p . The quantity m_e/m_p is related to α due to a coincidence in nuclear physics. These dependences are indicated explicitly in Fig. 1.

There are several amusing relationships between the different scales. For example, the size of a planet is the geometric mean of the size of the Universe and the size of an atom; the mass of man is the geometric mean of the mass of a planet and the mass of a proton. Such relationships, as well as the basic dependences on α and α_G from which they derive, might be regarded as coincidences if one did not appreciate that they can be deduced from known physical theory.

However, one of the scales in Fig. 1, that associated with the Universe, cannot be explained directly from known physics: it is apparently a coincidence that the present age of the Universe is of the order of α_G^{-1} times the electron timescale $\hbar/m_e c^2$. This led Dirac¹ to conjecture that G decreases with time as t^{-1} , so that the two timescales are always comparable. A more metaphysical explanation by Dicke² is that conditions are propitious for the existence of observers only when $t \approx \alpha_G^{-1} \hbar/m_e c^2$, so that this ‘coincidence’ should be of no surprise. This line of argument, which is discussed later, appeals to the ‘anthropic’ principle³. Later we shall also mention other features of the material world that seem sensitive to apparent coincidences among physical constants.

To describe structures on scales smaller than the atom one needs extra constants: in particular, the weak and strong inter-

action coupling constants g_w and f . It might be assumed that these are independent of α and α_G but there are some ‘anthropic’ interconnections between these numbers. For example, the condition that neutrinos can blow off the envelope of a star in its supernova phase will be shown to be, roughly speaking, $\alpha_G \sim \alpha_w^4$ where $\alpha_w = g_w^2 c/\hbar^3$ is the weak fine structure constant. Supernovae are essential if the heavy elements which are (presumably) necessary for life are to spread from their production sites throughout space. The same relationship between α_G and α_w explains why the cosmological helium production is $\sim 25\%$ by mass. If α_w were slightly smaller or larger, the helium production would be either 100% (in which case there would never be any water, perhaps another prerequisite for life) or 0% (in which case stellar evolution would be rather different). Finally there are coincidences between f and α and the elementary particle mass ratios which may be necessary for chemistry. There may be enough independent anthropic constraints to pin down the order of magnitude of α and α_G , and also that of α_w and f . Such considerations do not provide a real physical explanation, but they may indicate why these fundamental ratios are found to have their measured values.

One other important parameter in the Universe is the entropy per baryon or, equivalently, the photon-to-baryon ratio $\mathcal{S}$, which is of the order of 10^8 . The value of $\mathcal{S}$ has no explanation within the conventional hot big bang theory but it is also associated with a coincidence: if $\mathcal{S} \sim 10^8$, the matter and radiation densities are comparable at the time they thermally decouple. Although the anthropic principle does not seem to require a specific value for $\mathcal{S}$, it does require an upper limit on $\mathcal{S}$ of order $\alpha_G^{-1/4}$; otherwise, galaxies and stars would be unable to form through gravitational condensation. In fact, there are several schemes which suggest how an entropy parameter of the order of $\alpha_G^{-1/4}$ could be generated naturally.

Scales of structure in nature

We now present some simple arguments for the mass and length scales shown in Fig. 1. These are mostly straightforward consequences of simple physics. Many have been given before, but it is useful to bring them together. All our discussions will be based on order of magnitude arguments, so equations will be of the ‘ $\sim$ ’ rather than ‘ $=$ ’ kind, factors like π being neglected.

*Present address: California Institute of Technology, Pasadena, California 91109.

Protons: the size of a proton is taken to be the Compton wavelength associated with its rest mass:

$$r_p \sim \frac{\hbar}{m_p c} \sim 10^{-13} \text{ cm} \quad (1)$$

This is roughly the range of the strong interaction. Associated with r_p is the timescale $t_p = r_p/c \sim 10^{-23}$ s, the typical lifetime of a strong interaction resonance. By analogy with equation (1), the size or localisability of an electron is sometimes referred to as

$$r_e \sim \frac{\hbar}{m_e c} \sim 10^{-10} \text{ cm} \quad (2)$$

and we can define a corresponding electron timescale $t_e = r_e/c \sim 10^{-20}$ s.

Atoms: for a hydrogen atom the radius of the lowest energy electron orbit is

$$a_0 \sim \frac{\hbar^2}{m_e e^2} \sim 10^{-8} \text{ cm} \sim 1 \text{ Bohr} \quad (3)$$

and the associated energy is

$$E_0 \sim \frac{e^4 m_e}{\hbar^2} \sim 10 \text{ eV} \sim 1 \text{ Rydberg} \quad (4)$$

As one considers atoms of greater atomic number Z , the binding energy increases roughly as Z^2 and the radius of the inner shell electron orbit decreases roughly as Z^{-1} . However, the radius of the outer shell electron orbit remains of the order of a_0 , so all atoms have about the same size. If one introduces the fine structure constant

$$\alpha \equiv \frac{e^2}{\hbar c} \approx \frac{1}{137} \quad (5)$$

one can write a_0 and E_0 as

$$a_0 \sim \alpha^{-2} \left(\frac{e^2}{m_e c^2} \right) \quad E_0 \sim \alpha^2 m_e c^2 \quad (6)$$

The quantity $(e^2/m_e c^2)$ is the 'classical' radius of the electron $\sim 10^{-13}$ cm. Atomic masses span a range from 1 to roughly α^{-1} times m_p . The nuclei of atoms with atomic number exceeding α^{-1} have so much electrostatic binding energy that they are unstable to electron-positron pair production. In practice, the instability to fissioning comes in at a somewhat smaller atomic number.

The Planck scales: the only quantities of dimensions mass and length which can be constructed from G , $\hbar$ and c are the Planck scales:

$$M_{\text{Pl}} \sim \left(\frac{G}{\hbar c} \right)^{-1/2} \sim 10^{-5} \text{ g} \quad R_{\text{Pl}} \sim \left(\frac{G}{\hbar c^3} \right)^{1/2} \sim 10^{-33} \text{ cm} \quad (7)$$

By introducing the gravitational fine structure constant,

$$\alpha_G \equiv \frac{G m_p^2}{\hbar c} \approx 5 \times 10^{-39} \quad (8)$$

these scales can be expressed as

$$M_{\text{Pl}} \sim \alpha_G^{-1/2} m_p \quad R_{\text{Pl}} \sim \alpha_G^{1/2} r_p \quad (9)$$

So M_{Pl} is much larger than m_p but R_{Pl} is much smaller than r_p . The Planck length is the scale on which quantum gravitational fluctuations in the metric become of the order of unity, so the concept of space breaks down at such small scales. M_{Pl} can be interpreted as the mass of a black hole of radius R_{Pl} . Space may be thought of as being filled with virtual black holes of this size. Such 'instantons' have an important role in quantum gravity theory.

Black holes: the radius of a spherically symmetrical black hole of mass M is

$$R = \frac{2GM}{c^2} \sim \alpha_G \left(\frac{M}{m_p} \right) r_p \quad (10)$$

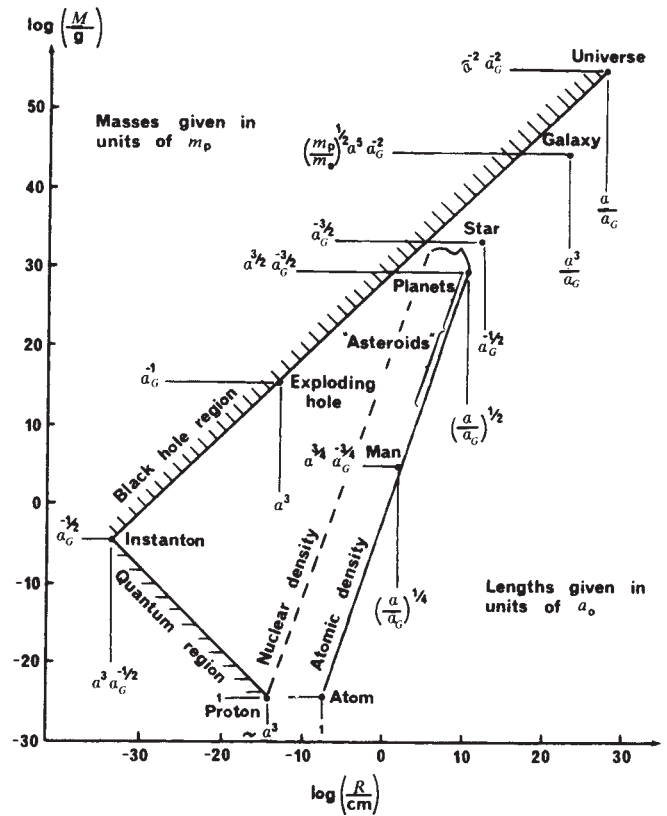


Fig. 1 The mass and length scales of various natural structures expressed in terms of the electromagnetic and gravitational fine structure constants, α and α_G . Some of these scales also depend on the electron-to-proton mass ratio, but this we have eliminated using $m_e/m_p \sim 10\alpha^2$. The asteroid scale also depends on the molecular weight A of rocky material. All these scales can be deduced directly from known physics except for the mass and length scale of the Universe, which depends on the age of the Universe being α_G^{-1} times the electron timescale $\hbar/m_e c^2$. Also shown are the atomic density line, the nuclear density line, the black hole density line and the 'quantum line' corresponding to the Compton wavelength. Most characteristic scales depend on simple powers of α_G ; the wide span of so many orders of magnitudes is a consequence of the huge numerical value of α_G^{-1} , which reflects the weakness of gravity on the microscopic scale.

This is the radius of the event horizon, the region from within which nothing can ever escape (at least, classically). Black holes larger than $1 M_\odot$ may form from the collapse of stars or dense star clusters. Smaller holes require much greater compression for their formation than could arise in the present epoch, but they might have been produced in the first instants after the big bang when the required compression could have occurred naturally. Such 'primordial' black holes could have any mass down to the Planck mass. In fact, Hawking⁴ has shown that small black holes are not black at all; because of quantum effects they emit particles like a black body of temperature given by

$$k\theta \sim \frac{\hbar c^3}{GM} \sim \alpha_G^{-1} \left(\frac{M}{m_p} \right)^{-1} m_p c^2 \quad (11)$$

This means that a hole of mass M will evaporate completely in a time

$$t_{\text{evap}} \sim \alpha_G^2 \left(\frac{M}{m_p} \right)^3 t_p (N(\theta))^{-1} \quad (12)$$

$N(\theta)$ is the number of species contributing to the thermal radiation: for $k\theta \ll m_e c^2$ these include only photons, neutrinos and gravitons; but at higher temperatures other species may contribute. The evaporation terminates in a violent explosion. For a solar mass hole, this quantum radiance is negligible: θ is

only 10^{-7} K and $t_{\text{evap}} \sim 10^{64}$ yr. But for small holes it is very important. A Planck mass hole has a temperature of 10^{32} K and only survives for a time $\sim R_{\text{Pl}}/c \sim 10^{-43}$ s. Those holes which are terminating their evaporation in the present epoch are particularly interesting. As the age of the Universe is $t_0 \sim \alpha_G^{-1} t_e$, the mass of such holes would be

$$M_h \sim \alpha_G^{-2/3} \left(\frac{t_0}{t_p} \right)^{1/3} m_p \sim \alpha_G^{-1} m_p \sim 10^{15} \text{ g} \quad (13)$$

and, from equation (10), their radius would be r_p . (The corresponding temperature is ~ 10 Mev: low enough to eliminate any uncertainty in $N(\theta)$ due to species of exotic heavy particles.)

Stars: Jordan⁵ first noted that the Sun has a mass of about $\alpha_G^{-3/2}$ times the proton mass; most stars have a mass in the range 0.1–10 times this. Jordan constructed an elaborate cosmological theory to explain this 'coincidence', but we will now see why stars are expected to lie in this mass range^{6,7}. The virial theorem implies that the gravitational binding energy of a star must be of the order of its internal energy. Its internal energy comprises the kinetic energy per particle (radiation pressure being assumed negligible for the moment) and the degeneracy energy per particle. The degeneracy energy will be associated primarily with the Fermi-momentum of the free electrons, $p \sim \hbar/d$, where d is their average separation. Provided the electrons are non-relativistic, the degeneracy energy is $p^2/2m_e$, so the virial theorem implies

$$kT + \frac{\hbar^2}{2m_e d^2} \sim \frac{GMm_p}{R} \sim \left(\frac{N}{N_0} \right)^{2/3} \frac{\hbar c}{d} \quad (14)$$

Here N is the number of protons in the star, $N_0 \equiv \alpha_G^{-3/2}$, and $R \sim N^{1/3} d$ is its radius. As a cloud collapses under gravity, equation (14) implies that, for large d , T increases as d^{-1} . For small d , however, T will reach a maximum

$$kT_{\text{max}} \sim \left(\frac{N}{N_0} \right)^{4/3} m_e c^2 \quad (15)$$

and then decrease, reaching zero when d is

$$d_{\text{min}} \sim \left(\frac{N}{N_0} \right)^{-2/3} r_e \quad (16)$$

A collapsing cloud becomes a star only if T_{max} is high enough for nuclear reactions to occur, that is $kT_{\text{max}} > qm_e c^2$ where q depends on the strong and electromagnetic interaction constants and is $\sim 10^{-2}$. From equation (15), one therefore needs $N > 0.1 N_0$. Once a star has ignited, further collapse will be postponed until it has burnt all its nuclear fuel. An upper limit to the mass of a star derives from the requirement that it should not be radiation-pressure dominated. Such a star would be unstable to pulsations which would probably result in its disruption. Using the virial theorem (that is, equation (14) with the degeneracy term assumed negligible) to relate a star's temperature T to its mass $M \sim Nm_p$ and radius R , the ratio of radiation pressure to matter pressure can be shown to be

$$\frac{P_{\text{rad}}}{P_{\text{mat}}} \sim \frac{aT^4 R^3}{NkT} \sim \left(\frac{N}{N_0} \right)^2 \quad (17)$$

so the upper limit to the mass of a star is also $\sim N_0 m_p$. A more careful calculation shows that there is an extra numerical factor of the order of 10, so one expects all main-sequence stars to lie in the range $0.1 \leq N/N_0 \leq 10$ observed.

In other words, only assemblages of 10^{56} – 10^{58} particles can turn into stable main sequence stars with H-burning cores. Less massive bodies held together by their own gravity can be supported by electron 'exclusion principle' forces at lower temperatures (they would not get hot enough to undergo nuclear fusion unless squeezed by an external pressure); heavier bodies are fragile and unstable owing to radiation pressure effects.

The M – R relationship for main-sequence stars is shown in Fig. 2, and can be deduced from the virial theorem and a knowledge of T , the central temperature. T adjusts itself so that

the nuclear energy generation rate balances the luminosity, the radiant energy content divided by the photon leakage time,

$$L \approx acT^4 R^4 / \kappa M \quad (18)$$

where κ is the opacity. The appropriate κ decreases as M increases (electron scattering being the dominant opacity for upper main-sequence stars) but the energy generation increases so steeply with T that M/R depends only weakly on M .

White dwarfs and neutron stars: when a star has burnt all its nuclear fuel, it will continue to collapse according to equation (14) and, providing it is not too large, it will end up as a cold electron-degeneracy supported 'white dwarf' with the radius $R \propto M^{-1/3}$ indicated by equation (16). However, when M gets so large that $kT_{\text{max}} \sim m_e c^2$, the electrons will end up relativistic, with the degeneracy energy being pc rather than $p^2/2m_e$. Thus the degeneracy term in equation (14) acts as d^{-1} instead of d^{-2} , and consequently there is no $T=0$ equilibrium state. From equation (15) this happens if M exceeds the mass

$$M_c \sim \alpha_G^{-3/2} m_p \sim 1 M_\odot \quad (19)$$

which characterises stars in general. A more precise expression for this critical value of M (Chandrasekhar mass) is $5.6 \mu^2 M_\odot$ where μ is the number of free electrons per nucleon⁸. For a star which has burned all the way up to iron, $\mu \approx 1/2$ and the limiting Chandrasekhar mass, taking into account the onset of inverse β -decay when electrons get relativistic, is $\sim 1.25 M_\odot$. Note that the white dwarf line in Fig. 2 meets the main-sequence line on the nuclear ignition ($T \sim 10^7$ K) isotherm. Stars bigger than M_c will collapse beyond the white dwarf density but may manage to shed some of their mass in a supernova explosion. The remnant core will comprise mainly neutrons, the electrons having been squeezed onto the protons through $p + e^- \rightarrow n + \nu$, and this core,

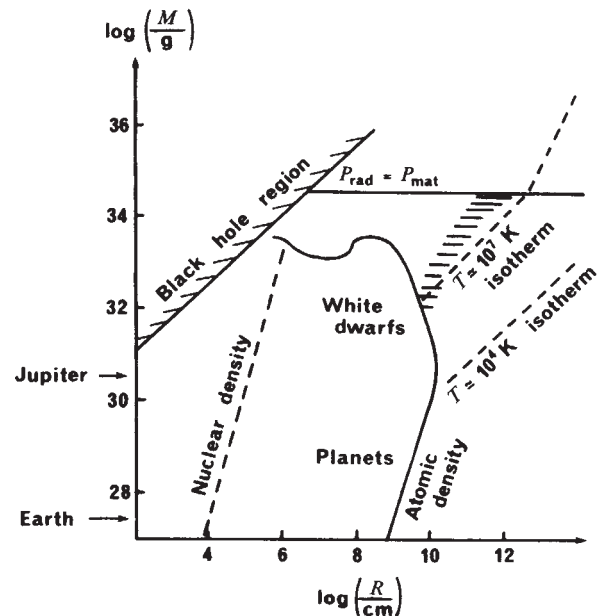


Fig. 2 An enlargement of part of Fig. 1. All solid planets lie on the atomic density line and all main sequence stars lie in the narrow mass range 0.1–10 times $\alpha_G^{-3/2} m_p$ (shaded). Regions with mass below this range would never get hot enough to ignite their nuclear fuel; regions with mass above this range would be radiation pressure dominated and consequently unstable. General relativistic instabilities are important for masses $\gg M_c$ (not shown). The precise form of the main-sequence band depends on the opacity and the nuclear reaction rate. All stars smaller than the Chandrasekhar mass ($\sim 1.4 M_\odot$) must eventually end up on the electron-degeneracy-supported white dwarf line. Stars bigger than this either shed some of their mass in a supernova explosion and end up as a neutron star or collapse to a black hole. Objects on the dotted line which bridges the white dwarf and nuclear density lines are unstable. Also shown are the ($T \sim 10^7$ K) nuclear ignition isotherm and the ($T \sim 10^4$ K) ionisation isotherm.

if small enough, may be supported by the neutron-degeneracy pressure. The limiting mass for a neutron star is more difficult to calculate than that for a white dwarf because of strong interaction effects and because, from equation (15), with $m_e \rightarrow m_p$, the particles which dominate the neutron star's mass are relativistic. The maximum mass is still of the order of $1 M_\odot$, however, and a neutron star bigger than this must collapse to a black hole. The maximum mass lies close to the intercept of the black hole line, given by equation (10), and the nuclear density line $\rho \sim m_p/r_p^3$. The intricacies of the line which bridges the white dwarf and neutron star regimes in Fig. 2 reflect the effects of gradual neutronisation and strong interactions. Stars on this bridge would be unstable and so are not of physical interest.

The above order-of-magnitude arguments show why the effects of radiation pressure, and relativistic degeneracy both become important for masses $\geq \alpha_G^{3/2} m_p$. (Note also that general relativity is unimportant for white dwarfs because the binding energy per unit mass is only $\sim (m_e/m_p)$ of c^2 at the Chandrasekhar limit.)

Planets and asteroids: the above analysis applies only to objects where the thermal energy kT and/or the degeneracy energy per electron exceeds the ionisation energy $\sim \alpha^2 m_e c^2$. If T is too small, the object will be solid or liquid rather than gaseous, in which case its structure is determined by the balance of degeneracy and chemical bonds rather than degeneracy and gravity. Since the degeneracy and electrostatic binding energy per particle are $\hbar^2 n^{2/3} m_e^{-1}$ and $e^2 n^{1/3}$, the density of any solid or liquid is of the order of the atomic density

$$\rho_0 \sim e^6 m_e^3 \hbar^{-6} \sim \frac{m_p}{a_0^3} \sim 1 \text{ g cm}^{-3} \quad (20)$$

From equation (16), the atomic density line meets the white dwarf line ($d_{\min} \sim a_0$) at a mass and radius

$$M \sim \left(\frac{\alpha}{\alpha_G}\right)^{3/2} m_p \sim \alpha^{3/2} M_c \sim 10^{30} \text{ g} \\ R \sim \left(\frac{\alpha}{\alpha_G}\right)^{1/2} a_0 \sim 10^{10} \text{ cm} \quad (21)$$

This point in Fig. 2 also lies on the $kT \sim \alpha^2 m_e c^2$ isotherm. Equation (21) characterises the maximum size of a planet and is of the order of the size of Jupiter. The central pressure of a self-gravitating body more massive than this can be balanced by degeneracy pressure only if the electrons are crushed closer together than in an ordinary solid. A lower limit for the size of a (solid) planet can be specified if the planet is reasonably round—that it be bigger than its own mountains. The maximum size of a mountain can be found from the following argument⁷. A mountain of height h must not provide so much pressure on the planet's surface that it liquifies its base. The condition for this is easily shown to be

$$h < h_{\max} \sim \frac{E_{\text{liq}}}{A m_p g} \quad (22)$$

where A is the molecular weight of the planetary material, g is the surface gravity and E_{liq} is the liquefaction energy per molecule. The binding energy of a solid is about 0.1 Rydberg per molecule and E_{liq} is about a tenth of this (as only the directionality of the bonds is broken in liquefaction), so

$$E_{\text{liq}} \sim \beta \alpha^2 m_e c^2 \quad (\beta \sim 10^{-2}) \quad (23)$$

For the Earth, $g \sim 10^3 \text{ cm s}^{-2}$ and equation (22) implies a maximum height of the order of 10 km. In general one has

$$g \sim \frac{GM}{R^2} \sim G \left(\frac{M}{m_p}\right)^{1/3} A^{2/3} (y a_0)^{-2} m_p \quad (24)$$

where $y a_0$ is the molecular size of the planetary material. (A solid terrestrial-type planet is mostly SiO_2 and iron, for which $A \sim 60$ and $y \sim 4$.) Equations (22–24) thus imply

$$h_{\max} \sim (\beta y^2) \left(\frac{\alpha}{\alpha_G}\right) \left(\frac{M}{m_p}\right)^{-1/3} A^{-5/3} \quad (25)$$

This is less than the radius of the planet, $R \sim y a_0 (M/A m_p)^{1/3}$, provided M and R exceed

$$M \sim (y\beta)^{3/2} A^{-2} \left(\frac{\alpha}{\alpha_G}\right)^{3/2} m_p, \quad R \sim A^{-1} \left(\frac{\alpha}{\alpha_G}\right)^{1/2} a_0 (y^3 \beta)^{1/2} \quad (26)$$

which is thus the maximum size of an irregularly shaped 'asteroid'. The value of A is constrained by the requirement that solid planets must be made of material whose atomic number is sufficiently high that it is not vaporised by the high temperatures attained during the formation of the Solar System.

'Habitable' planets: the mass range of habitable planets can be narrowed down still further if an atmosphere and an appropriate surface temperature are required⁹. The optimum temperature for organisms is $T \sim \epsilon/k$, where $\epsilon \sim 10^{-3}$ Rydberg is the typical energy released in relevant reactions involving complex molecules: were T much larger, the molecules would be disrupted; were it much smaller, biological processes would require an exponentially long timescale. If an atmosphere composed of gas heavier than hydrogen were necessary for life, then the thermal velocity of hydrogen at this temperature should be slightly greater than the escape velocity from the planet, which implies

$$\frac{GMm_p}{R} \sim \epsilon \alpha^2 m_e c^2 \quad (27)$$

With the relationship $M/R^3 \sim A m_p / y^3 a_0^3 \sim m_p / a_0^3$, this implies that a life-supporting planet must have a mass

$$M \sim \epsilon^{3/2} \left(\frac{\alpha}{\alpha_G}\right)^{3/2} m_p \quad (28)$$

Only a small fraction of planets in the mass range permitted by equations (21) and (26) would satisfy this criterion.

Press⁹ has used this line of argument to set an upper limit to the size of living creatures on the basis that they must not 'break' when they fall: the energy released by a body of mass M_m in falling from height h_m on a planet whose surface gravity is given by equation (28) must be insufficient to break the molecular bonds on an area h_m^2 . This yields an upper limit

$$M_m h_m \left(\frac{GM}{R^2}\right) \leq \left(\frac{M_m}{m_p}\right)^{2/3} \epsilon \alpha^2 m_e c^2 \quad (29)$$

Substituting for M with equation (28) and assuming atomic density, this implies

$$h_m \sim \epsilon^{1/4} \left(\frac{\alpha}{\alpha_G}\right)^{1/4} a_0 \sim \left(\frac{\epsilon}{0.001}\right)^{1/4} \text{ cm} \quad (30)$$

with only a weak dependence on ϵ . If one assumes that the surface stresses are distributed along faults as wide as a polymer, the estimate for h_m is increased by a factor of 100. (In fact, equation (30) is more appropriate as the estimate for the size of a water droplet, which can be derived from a similar argument⁷.) Allowing for the factor of 100, the mass associated with h_m is

$$M_m \sim 10^3 \left(\frac{\alpha}{\alpha_G}\right)^{3/4} m_p \sim 10^5 \text{ g} \quad (31)$$

which is roughly the mass of a man. We plot this mass in Fig. 1, although its value is very uncertain (and is in any case irrelevant to swimming organisms); note its dependence on $\alpha_G^{3/4}$.

The Universe: in the simplest Friedmann cosmological model, the age of the Universe t_0 is of the order of H_0^{-1} where H_0 is the Hubble parameter. (This relation fails only if the Universe is closed and near its maximum expansion.) Since $H_0 \sim 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$, this implies $t_0 \sim 10^{10} \text{ yr}$, a conclusion which is supported by several independent arguments. The associated horizon size (the distance travelled by light since the beginning of the Universe) satisfies the approximate relationship

$$ct_0 \sim \alpha_G^{-1} \left(\frac{\hbar}{m_e c}\right) \sim \left(\frac{\alpha}{\alpha_G}\right) a_0 \quad (32)$$

In other words the ratio of the size of the observable Universe to the size of an atom is comparable with the ratio of the electrical (or nuclear) and gravitational forces between elementary particles. There is no explanation for this well known coincidence within conventional physics, but Dirac¹ has conjectured that α_G is always given by

$$\alpha_G \sim \frac{\hbar}{m_e c^2 t} \sim \left(\frac{t}{t_e}\right)^{-1} \quad (33)$$

Assuming that $\hbar$, c , and m_e are constant in time, this requires that G decreases as t^{-1} , so Dirac invokes equation (32) as the basis for a new cosmology. Such a variation of G may be inconsistent with observation.

An alternative 'anthropic' explanation for equation (32) was first suggested by Dicke². As life presumably requires elements heavier than hydrogen and helium and such elements can only be produced and spread throughout the Universe by supernovae, Dicke argued that any 'cognisable' Universe would be one in which some stars have already completed their main-sequence evolution. The luminosity of a star whose opacity is dominated by electron scattering (as applies for large stars) is $\sim (P_{\text{rad}}/p)L_E$; (P_{rad}/p) is given in terms of M by equation (17) and L_E is the Eddington luminosity, $4\pi GM_p c/\sigma_T$, where $\sigma_T \approx \alpha^2 r_e^2$ is the Thomson cross-section. A characteristic timescale¹⁰ is that over which an object of luminosity L_E would radiate away its entire rest mass:

$$t_s = \frac{c\sigma_T}{4\pi Gm_p} = \left(\frac{\alpha^2}{\alpha_G}\right) \left(\frac{m_p}{m_e}\right)^2 t_p \quad (34)$$

and this timescale involves the 'large number' α_G^{-1} explicitly. If $\eta \approx 10^{-2}$ is the fraction of a star's rest mass that can be released through nuclear burning, the main sequence lifetime is thus

$$t_{\text{MS}} \sim \eta \left(\frac{P_{\text{rad}}}{p}\right)^{-1} t_s \approx \left\{ \eta \alpha^2 \left(\frac{m_p}{m_e}\right)^2 \right\} \alpha_G^{-1} \left(\frac{M}{M_c}\right)^{-2} t_p \quad (35)$$

The quantity in braces is of the order of unity; as stars must have a mass of the order of M_c , living observers could exist only when

$$t_0 > t_{\text{MS}} \sim \alpha_G^{-1} t_p \quad (36)$$

However, t_0 cannot be much bigger than t_{MS} or else most of the Universe would have been processed into white dwarfs, neutron stars or black holes. Therefore observers are most likely to exist at an epoch $t_0 \sim t_{\text{MS}}$, and they will find equation (32) automatically fulfilled.

In fact, t_0 exceeds the value of t_{MS} in equation (36) by a factor $m_p/m_e \sim 10^3$. However, t_{MS} is sensitive to the value of M and equation (36) is really appropriate only for an upper main-sequence star. Dicke's argument would therefore be more convincing if one could show that the first stars would be of low mass. A possible reason might come from the general argument^{11,12} that the mass scale at which a collapsing cloud must stop fragmenting is

$$M_{\text{frag}} \sim x^{-1/2} \left(\frac{kT}{m_p c^2}\right)^{1/4} M_c \sim 10^{-2} x^{-1/2} M_c \quad (37)$$

where T is the cloud's temperature (expected to be $\approx 10^4$ °K) and x is the ratio of its luminosity to that of a black body with this temperature. M_{frag} has only a weak dependence on T and is at least an order of magnitude smaller than M_c , thus increasing t_{MS} to a value more comparable to t_0 . Alternatively, one might argue that the first stars do have $t_{\text{MS}} \ll t_0$ but that some of the elements vital for life must be generated through the s-processes associated with later-forming less massive stars.

The total mass associated with the observable Universe (the mass within the horizon volume) is $\sim \rho_0 c^3 t_0^3$, where ρ_0 is the

present matter density, given by the Friedmann equation:

$$\rho_0 = \frac{3H_0^2}{8\pi G} + \frac{Kc^2}{16\pi G} \quad (38)$$

Here K is the scalar curvature of the Universe. Providing the K term is smaller than the others, we deduce that the mass of the Universe is

$$M_u \sim c^3 t_0^3 G^{-1} H_0^2 \sim \frac{c^3 t_0}{G} \sim \alpha_G^{-2} \left(\frac{m_p}{m_e}\right) m_p \quad (39)$$

The fact that the number of protons in the Universe is of the order of α_G^{-2} , another cosmological coincidence, is thus explained providing one can justify neglecting the K term in equation (38). It has been argued that K must always be zero by appealing to Mach's principle¹³ but, apart from this, there may be anthropic reasons for expecting that the K term is small. If K is negative, galaxies could not have condensed out from the general expansion unless $(-K)$ were less than $G\rho/c^2$ at their formation epoch. Otherwise, their gravitational binding energy would not have been large enough to halt their expansion. $(-K)$ may, in fact, exceed $G\rho/c^2$ now, but not by a large factor. If K is positive, it must be $< G\rho_0/c^2$, otherwise the whole Universe would have recollapsed before $t \approx t_{\text{MS}}$. The simple relationship between M_u and t_0 given by equation (39), with equation (36), explains why the Universe must be as big and diffuse as it is to last long enough to give rise to life. Relationships (32) and (39) also mean that the Universe has an optical depth of the order of unity to electron scattering.

Galaxies: it is not certain how galaxies form, so any estimate of their scale is very model dependent. The scale indicated in Fig. 1 derives from the following argument¹⁴⁻¹⁶. We assume that galaxies originate from overdense regions in the gaseous primordial material, and that they have a mass M and radius R_B when they become bound. After binding, protogalaxies may virialise at a radius $\sim R_B/2$. Thereafter, they will deflate on a cooling timescale, with a virial temperature

$$T \sim \frac{GM_p}{kR} \quad (40)$$

Providing kT exceeds one Rydberg the dominant cooling mechanism is bremsstrahlung and the associated cooling timescale is

$$t_{\text{cool}} \sim \frac{1}{n\alpha\sigma_T c} \left(\frac{kT}{m_e c^2}\right)^{1/2} \approx \frac{m_e^2 c^3}{\alpha e^4 n} \left(\frac{kT}{m_e c^2}\right)^{1/2} \quad (41)$$

The free-fall timescale is

$$t_{\text{ff}} \sim \left(\frac{GM}{R^3}\right)^{-1/2} \quad (42)$$

and this exceeds t_{cool} when R falls below a mass-independent value

$$R_g \sim \alpha^4 \alpha_G^{-1} \left(\frac{m_p}{m_e}\right)^{1/2} a_0 \quad (43)$$

which, from a more precise calculation, is 75 kpc. Until a massive cloud gets within this radius it will contract quasistatically and cannot fragment into stars.

This argument applies only if the mass is so high that $kT_{\text{virial}} \gg \alpha^2 m_e c^2$ at the 'magic radius' R_g : that is,

$$M \gg M_g = \alpha_g^{-2} \alpha^5 \left(\frac{m_p}{m_e}\right)^{1/2} \approx 10^{12} M_\odot \quad (44)$$

Gas clouds of mass $< M_g$ cool more efficiently owing to recombination, and can never be pressure supported. Thus, M_g is a characteristic maximum galactic mass. Primordial clouds of mass $> M_g$ are inhibited from fragmentation and may remain as

hot pressure-supported clouds. This type of argument can be elaborated and made more realistic¹⁴⁻¹⁶; but one still obtains a mass $\sim M_g$ above which any fluctuations are likely to remain amorphous and gaseous, and which may thus relate to the mass-scale of galaxies. M_g and R_g may thus characterise the mass and radius of a galaxy. Of all the scale estimates in Fig. 1, this is the least certain—the properties of galaxies may be a consequence of irregularities imprinted in the Universe by processes at early epochs.

This completes the justification of Fig. 1 except that we may eliminate the dependences on m_e/m_p using the relationship

$$m_e/m_p \sim 10\alpha^2 \quad (45)$$

which results from a coincidence in the nuclear physics (see equation (58)). From Fig. 1 some amusing interconnections can be deduced. Regarding the length scales, for example; $\text{man} \sim (\text{planet} \times \text{atom})^{1/2}$; $\text{planet} \sim (\text{Universe} \times \text{atom})^{1/2}$. Regarding the mass-scales: $\text{Planck} \sim (\text{exploding hole} \times \text{proton})^{1/2}$; $\text{exploding black hole} \sim (\text{Universe} \times \text{proton})^{1/2}$; $\text{man} \sim (\text{planet} \times \text{proton})^{1/2}$. These relationships cannot of course be regarded as coincidences as they can be expected *a priori*.

Other cosmological coincidences

Another dimensionless constant which characterises our Universe is the photon-to-baryon ratio $\mathcal{S} \sim 10^8 \Omega^{-1} g^{-1}$ (where Ω is the baryon density in units of the critical density). According to the hot big bang model the background radiation dominated the density of the Universe until a time

$$t_{\text{eq}} \sim \mathcal{S}^2 \alpha_G^{-1/2} t_p \sim 10^{11} \Omega^{-2} \text{ s} \quad (46)$$

and thermally decoupled from the matter when T fell below $\sim 0.1\alpha^2 m_e c^2/k$ at a later time

$$t_{\text{dec}} \sim \mathcal{S}^{1/2} \alpha_G^{-1/2} \alpha^{-3} \left(\frac{m_p}{m_e}\right)^{3/2} t_p \sim 10 \text{ g}^3 \Omega^{-1/2} \text{ s} \quad (47)$$

(These equations derive from the fact that the radiation density and temperature fall like R^{-4} and R^{-1} respectively. Here R is the length scale of the Universe, which is $\propto t^{1/2}$ for $t < t_{\text{eq}}$ and $\propto t^{2/3}$ for $t > t_{\text{eq}}$.) The observation that t_{eq} and t_{dec} are within an order of magnitude of each other (for $\Omega \sim 0.1$) corresponds to the coincidence

$$\mathcal{S} \sim 10\alpha^{-2} \left(\frac{m_p}{m_e}\right) \sim \alpha^{-4} \quad (48)$$

Now, the formation of galaxies cannot occur until, first, T has fallen below the decoupling temperature, and second, the radiation density has fallen below the matter density. Thus, the anthropic principle requires that t_0 exceeds both t_{eq} and t_{dec} , these conditions being satisfied if

$$\mathcal{S} < \alpha_G^{-1/4} \left(\frac{m_p}{m_e}\right)^{1/2} \sim \alpha_G^{-1/4} \alpha^{-1} \sim 10^{11} \quad (49)$$

A lower limit on $\mathcal{S}$ is obtained if one requires that the Jeans mass in the period t_{eq} to t_{dec} ,

$$M_J \sim G^{-3/2} p^{3/2} \rho^{-2} c^3 \sim G^{-1} t_{\text{eq}} c^3 \sim \alpha_G^{-3/2} \mathcal{S}^2 m_p \quad (50)$$

exceed the galactic mass indicated by equation (44). Such a condition would have to be satisfied if one identified M_J with the mass of a supercluster of galaxies¹² and it implies

$$\mathcal{S} > \alpha_G^{-1/4} \alpha^{9/2} \left(\frac{m_p}{m_e}\right)^{5/4} \sim \alpha_G^{-1/4} \alpha^2 \sim 10^6 \quad (51)$$

Equations (49) and (51) constrain $\mathcal{S}$ to lie fairly close to its observed value.

The rough equality between t_{eq} and t_{dec} has prompted suggestions that the background radiation may have been generated in some way during the early history of the Universe. For example, the radiation may have been produced by a first generation of pregalactic objects¹⁷. Such objects have a characteristic

lifetime $\sim t_{\text{MS}}$ given by equation (35), so after t_{MS} the radiation density would be expected to be ηF times the matter density, where F is the fraction of the Universe which goes into the stars and η is the fraction of each star's rest mass released through nuclear energy generation and supernova outbursts. The value of t_{eq} associated with the generated $\mathcal{S}$ is thus $(\eta F)^{3/2} t_{\text{MS}}$ and equations (35) and (46) imply, for objects with $L \approx L_E$, that

$$\mathcal{S} \sim \alpha_G^{-1/4} \left\{ \alpha \left(\frac{m_p}{m_e}\right) F^{3/4} \eta^{5/4} \right\} \quad (52)$$

If the term in braces is of the order of unity, one would expect $\mathcal{S} \sim \alpha_G^{-1/4}$ as observed. The same relationship between $\mathcal{S}$ and α_G would apply if the radiation were generated by black holes accreting at the Eddington limit¹⁸. Another possibility^{19,20} is that a photon-baryon ratio of the order of

$$\mathcal{S} \sim \left(\frac{G m_w^2}{\hbar c}\right)^{1/4} \quad (53)$$

where m_w is the W-boson mass, could be generated by CP violating processes in primordial black hole evaporation, or by the large fluctuations which might arise at the epoch of weak/electromagnetic interaction symmetry breaking¹⁸. This is just the relation $\mathcal{S} \sim \alpha_G^{-1/4}$ with m_p replaced by m_w . One could probably conceive of several cosmological scenarios in which a value of $\mathcal{S}$ of the order of $(G m^2/\hbar c)^{-1/4}$ would be generated, where m is the mass of some elementary particle. The important point is that the coincidences involving $\mathcal{S}$ can be explained naturally without recourse to anthropic considerations. (The processes of primordial nucleosynthesis, apart from deuterium production, are insensitive to $\mathcal{S}$ provided that $\mathcal{S} \geq 10^3$.)

Dicke's anthropic interpretation² of equation (32) was advanced before the microwave background was discovered. One could perhaps replace (or strengthen) the argument leading to equation (36) by the more general statement—independent of considerations of stellar physics—that observers require $kT \leq 0.1\alpha^2 m_e c^2$ and the possibility of thermodynamic disequilibrium, and therefore could survey the Universe only when $t_0 \geq t_{\text{dec}}$. From equation (46), if $\mathcal{S} \sim \alpha_G^{-1/4}$ the condition $t_0 > t_{\text{eq}}$ is the same as equation (36).

Dicke's argument helps us to understand why the 'coincidence' of equation (32) prevails—why a similar large number arises in two contexts that might at first sight seem unrelated. It does not tell us why α_G^{-1} has its particular immense value. The task of deriving such a large pure number from basic theory might seem a daunting one. However, Zeldovich²¹ points out that a complete quantum theory of gravity, where tunnelling and topology changes play a key role, is likely to involve exponentials of pure numbers, so that a factor $\sim 10^{40}$ may readily arise. Although no such number has been predicted, a relationship between α and α_G may be forthcoming from quantum field theory. It has been suggested^{22,23} that all space integrals in quantum electrodynamics should be cut off at the Planck length (given by equation (7)), thereby reducing otherwise divergent integrals to finite functions of the parameter $(\alpha \log \alpha_G)$. Various arguments suggest that a self-consistent electrodynamics is possible only if this parameter has some specific value of the order of unity, that is, one requires

$$\alpha^{-1} \sim \log \alpha_G^{-1} \quad (54)$$

This argument may not be convincing, but empirically relation (54) is satisfied. Figure 1 mainly shows that all objects where gravity is important have masses exceeding m_p by some simple power of α_G^{-1} . As the scaling is so simple, could one envisage a hypothetical universe in which all microphysical laws were unchanged, but G was (say) a million times stronger? Planetary and stellar masses ($\propto \alpha_G^{-3/2}$) would then be lowered by 10^9 ; but hydrogen-burning main-sequence stars would still exist albeit with lifetimes ($t_{\text{MS}} \propto \alpha_G^{-1}$ according to equation (35)) of only ~ 100 yr rather than $\sim 10^8$ yr. Moreover, Dicke's argument would still apply: a hypothetical observer looking at the

Universe when $t_0 \approx t_{MS}$ would still find equation (32) fulfilled, although the number of particles in his universe would be 10^{12} times lower than in ours. Are there anthropic arguments against the 'cognisability' of the kind of small-scale speeded-up universe that corresponds to a much smaller α_G^{-1} ?

One might argue that the gradual evolutionary emergence of complex organisms demands a large ratio between cosmic and microphysical timescales, and requires also that organised structures can become large (masses $\gg m_p$) before gravity overwhelms chemical forces. This favours a very large α_G^{-1} ; moreover, as $M_u/M_c \approx \alpha_G^{-1}$, a universe where α_G^{-1} is large contains more independent sites where evolution might occur. But there is no basis here for being at all quantitative. There are, however, some more specific (though somewhat contrived) anthropic arguments that pin down α_G .

One connection between α and α_G , given by Carter^{3,24}, is related to the existence of convective stars. If radiation transport is unable to maintain a star's surface temperature T_s above the ionisation temperature $\sim 0.1\alpha^2 m_e c^2/k$, a convective outer layer develops and this supplements the heat transport so that it does. The value of T_s which can be maintained by radiative transport, assuming (from the Coulomb penetration condition) a central temperature $T_c \sim 10^{-2}\alpha^2 m_p a^2/k$, is

$$T_s \sim \left(\frac{L}{aR^2}\right)^{1/4} \sim (\text{optical depth through star})^{-1/4} \cdot T_c \quad (55)$$

T_s exceeds $0.1\alpha^2 m_e c^2/k$ provided M exceeds

$$M_* \sim M_c \alpha_G^{-1/2} \alpha^6 \left(\frac{m_e}{m_p}\right)^2 \sim M_c \alpha_G^{-1/2} \alpha^{10} \quad (56)$$

using relation (45). The mass M_* thus divides the (convective) red dwarfs from the (radiative) blue giants. This lies in the range around M_c in which main-sequence stars exist only because

$$\alpha_G \sim \alpha^{20} \quad (57)$$

Were G (and hence α_G) slightly larger, all stars would be blue giants; if it were slightly smaller, all stars would be red dwarfs. Carter ascribes anthropic significance to relation (57) on the basis that the formation of planetary systems may be associated with convective stars. (This is supported by the observation that red dwarfs have much less angular momentum than blue giants and a loss of angular momentum may be a consequence of planet formation, but this is not a compelling argument. Alternatively, convective stars may lose their angular momentum through a wind, and this wind, by blowing away the gaseous envelope of planets close to the star, facilitates the formation of solid planets with non-hydrogen atmospheres.) Carter infers that no planets, and hence no life, would form if α_G were much larger than α^{20} . If α_G were much smaller, all stars would be chemically homogeneous due to convective mixing and one would not get the 'onion-skin' shell structure which characterises pre-supernova models. Carter's convective condition^{12,14} is also the condition that the number of stars in a galaxy be the same as the number of galaxies in the Universe.

The constants pertaining to nuclear physics are the strong (scalar) coupling constant $f^2 \approx 15$ and the mass ratios $m_e/m_N \approx 1/1837$, $m_\pi/m_N \approx 1/7$ and $\Delta_N/m_N \approx 1/730$ where $\Delta_N = (m_N - m_p)$. (Many more mass ratios are associated with high-energy physics but it is not clear which of them are fundamental.) The important features of nuclear physics depend²⁴ on the following four coincidences:

$$\begin{array}{llll} f^2 \approx 2m_N/m_\pi & \Delta_N/m_e \approx 2 & \alpha \approx \Delta_N/m_\pi & f \approx 1/3\alpha^{1/2} \\ (a) & (b) & (c) & (d) \end{array} \quad (58)$$

Equation (58a) implies that strong interactions are only marginally strong enough to bind nucleons into nuclei. If f were

slightly weaker, only hydrogen could exist; if it were slightly stronger nuclei of almost unlimited size might exist. (58b) implies that neutrons are unstable to β -decay in isolation but not in the presence of relativistic degenerate electrons. (58c) implies that the electrostatic energy in light nuclei $\sim \alpha m_\pi$ is comparable to the neutron-proton mass difference. (58d) implies that the electrostatic energy is small compared with nuclear binding energy in light nuclei but comparable to it for nuclei with $Z \sim (f\alpha)^{-3} \sim 30$, so such large nuclei are unstable to electromagnetic disruption. If the relations indicated by equation (58) were not satisfied, elements vital to life would not exist, so one might also ascribe anthropic significance to these relations. Note that the combination of (58a), (b), (c) and (d) implies equation (45). Kahn²⁷ noted that $m_e \ll m_p$ may be a prerequisite for complex chemistry, as this ensures that the ions are located to a precision $\sim (m_e/m_p)^{1/4}$ times their mean spacing.

There may be other anthropic aspects to the value of f . For example, elements heavier than helium can form in stars through the triple- α reaction ($\text{He}^4 + \text{He}^4 \rightarrow \text{Be}^8$, $\text{Be}^8 + \text{He}^4 \rightarrow \text{C}^{12}$), and it is apparently accidental that this process yields a substantial amount of carbon^{25,26}. ^8Be is unstable (otherwise the 'helium flash' in giants would lead to a catastrophic explosion); but the reaction proceeds to ^{12}C because this latter nucleus happens to have an energy level just above the sum of the energies of ^8Be and ^4He . There is, however, no similar favourably placed resonance in ^{16}O ; otherwise almost all the carbon would be transmuted into oxygen. If f were even merely a few per cent larger, there would be 100% cosmological helium production because He^2 would be bound, and deuterium could form by $p + p \rightarrow ^2\text{He} + \gamma$, $^2\text{He} \rightarrow \text{D} + e^+ + \nu$, even if there were no frozen-out neutrons. A small increase in f would also affect the binding energies of heavier nuclei.

Few features of the physical world seem to depend on the actual value of the weak interaction coupling constant. There are, however, two processes, both crucial to nucleosynthesis, which are sensitive to neutrino cross-sections and thus imply a coincidence of 'anthropic' relevance.

The first of these is connected with the production of helium through cosmological nucleosynthesis. The helium abundance Y is essentially determined by the temperature T_F at which the neutron-proton ratio freezes out. This occurs when the weak interactions ($p + e^- \rightarrow n + \nu$, $p + \bar{\nu} \rightarrow n + e^+$), proceeding at a rate $\propto g_w^2 T^5$, where $g_w \sim 10^{-49} \text{ erg cm}^3$ is the weak coupling constant, become slower than the cosmological expansion rate $\sim (GaT^4/c^2)^{1/2}$ at a temperature

$$kT_F \sim G^{1/6} g_w^{-2/3} \hbar^{11/6} c^{-5/6} \quad (59)$$

As virtually all the frozen-out neutrons burn into helium, the resultant helium abundance is

$$Y \sim \frac{2n_N/n_p}{n_N/n_p + 1} \quad \text{where} \quad \frac{n_N}{n_p} \sim \exp\left(\frac{-\Delta_N c^2}{kT_F}\right) \quad (60)$$

The fact that $Y \sim 25\%$ rather than 0% or 100% results only because $kT_F \sim \Delta_N c^2 \sim m_e c^2$ and, from equation (59), this derives from the 'coincidence'

$$\left(\frac{Gm_e^2}{\hbar c}\right)^{1/4} \sim \left(\frac{g_w m_e^2 c}{\hbar^3}\right) \quad (61)$$

The dimensionless number on the right may be regarded as the weak fine structure constant ($\alpha_w \sim 10^{-11}$) and the number on the left as the quarter power of the electron gravitational fine structure constant. In the Weinberg-Salam unified theory of weak and electromagnetic interactions²⁸ α_w is related to α by

$$\alpha_w \sim \alpha \left(\frac{m_e}{m_w}\right)^2 \quad (62)$$

So equation (61) combined with equation (45) may be written as

$$\alpha_G \sim \left(\frac{m_e}{m_W}\right)^8 \left(\frac{\alpha^2 m_p}{m_e}\right)^2 \sim \left(\frac{m_e}{m_W}\right)^8 \quad (63)$$

It is unclear to what extent these coincidences can be interpreted anthropically. Relation (63), for example, may reflect some deep-rooted connection between gravity and the weak and electromagnetic interactions. However, life could probably not exist if Y were 100% (as would be the case if α_W were slightly smaller), as there would be no water. Also, the lifetime of a helium star is less than that of a hydrogen star and might not be long enough to permit the evolution of life. It is not obvious that 0% helium production (as would arise if α_W were slightly larger) would be incompatible with life.

Equation (61) may, however, be associated with another anthropic condition (which would limit α_W in both directions)—the existence of supernovae. What actually ejects the stellar envelope in a supernova explosion is still uncertain, but it may be the outward surge of neutrinos generated by the high temperatures in the collapsing core³⁰. For this model to work, one requires the timescale on which neutrinos interact with nuclei in the envelope to be comparable to the dynamical timescale. If it were much longer, the envelope would be essentially transparent to the neutrinos; if it were shorter, the neutrinos would be trapped in the core, and could not escape to deposit their momentum in the less tightly bound surrounding layers. The two timescales are comparable if

$$c^{-3} \hbar^{-4} g_W^2 n (kT)^2 \sim (G n m_p)^{1/2} \quad (64)$$

where n is the nucleon number density. For the neutrinos to be produced at all (by $e^+ + e^- \rightarrow \nu + \bar{\nu}$) one requires kT to be of the order of $m_e c^2$. One also expects the density at the bounce to be of the order of the nucleon degeneracy density which, for $M \sim M_\odot$, is of the order of the nuclear density: $n \sim (\hbar/m_p c)^{-3}$. Putting these values for n and T into equation (64) gives relation (61) apart from a $(m_e/m_p)^{1/2}$ factor. Hence, the condition that stars go through a supernova phase is essentially the same as the condition that there be an interesting amount of cosmological helium production.

Conclusion

The possibility of life as we know it evolving in the Universe depends on the values of a few basic physical constants—and is in some respects remarkably sensitive to their numerical values. Indeed, the various anthropic relations quoted above in principle determine the order-of-magnitude of most of the fundamental constants of physics. Equations (54) and (57) specify α and α_G ; equations (61) and (58d) specify α_W and f ; and equations (9), (45), (58) and (63) determine m_p , m_e , m_μ , m_n and m_W in terms of the Planck mass. From a physical point of view, the anthropic 'explanation' of the various coincidences in nature is unsatisfactory, in three respects. First, it is entirely *post hoc*; it

has not yet been used to predict any feature of the Universe (although some people have used it to rule out various cosmological models). Second, the principle is based on what may be an unduly anthropocentric concept of an observer. The arguments invoked here assume that life requires elements heavier than hydrogen and helium, water, galaxies, and special types of stars and planets. It is conceivable that some form of intelligence could exist without all of these features—thermodynamic disequilibrium is perhaps the only prerequisite that we can demand with real conviction. Third, the anthropic principle does not explain the exact values of the various coupling constants and mass-ratios, only their order of magnitudes. With enough anthropic conditions, one may be able to be more precise about the constants of nature, but the present situation is unsatisfactory.

On the other hand, nature does exhibit remarkable coincidences and these do warrant some explanation. Apart from Dirac's suggestion¹ (which only considers the age of the Universe coincidence), the anthropic explanation is the only candidate and the discovery of every extra anthropic coincidence increases the *post hoc* evidence for it. The concept would be more palatable if it could be given a more physical foundation. Such a foundation may already exist in the Everett 'many worlds' interpretation of quantum mechanics, according to which, at each observation, the Universe branches into a number of parallel universes, each corresponding to a possible outcome of the observation³⁰. The Everett picture is entirely consistent with conventional quantum mechanics; it merely bestows on it a more philosophically satisfying interpretation. There may already be room for the anthropic principle in this picture. Wheeler³¹ envisages an infinite ensemble of universes, all with different coupling constants and so on. Most are 'still-born', in that the prevailing physical laws do not allow anything interesting to happen in them; only those which start off with the right constants can ever become 'aware of themselves'. One would have achieved something if one could show that any cognisable universe had to possess some features in common with our Universe. Such an ensemble of universes could exist in the same sort of space as the Everett picture invokes. Alternatively, an observer may be required to 'collapse' the wave function³². These arguments go a little way towards giving the anthropic principle the status of a physical theory but only a little: it may never aspire to being much more than a philosophical curiosity. One day, we may have a more physical explanation for some of the relationships discussed here that now seem genuine coincidences. For example, the coincidence $\alpha_G \sim (m_e/m_W)^8$, which is essential for nucleogenesis, may eventually be subsumed as a consequence of some presently unformulated unified physical theory. However, even if all apparently anthropic coincidences could be explained in this way, it would still be remarkable that the relationships dictated by physical theory happened also to be those propitious for life.

We thank John Barrow, Dick Bond, Bill DeCampi, Max Delbrück, Carol Readhead, Vernon Sandberg, Frank Tipler and Steve Webber for helpful discussions. B.J.C. is grateful to the Lindemann Trust and the NSF for financial support.

1. Dirac, P. A. M. *Nature* **139**, 323 (1937); *Proc. R. Soc.* **165A**, 199 (1938).
2. Dicke, R. H. *Nature* **192**, 440 (1961).
3. Carter, B. in *Confrontation of Cosmological Theories with Observation* (ed. Longair, M. S.) 291 (Reidel, Dordrecht, 1974).
4. Hawking, S. W. *Nature* **248**, 30 (1974).
5. Jordan, P. *Die Herkunft der Sterne* (Stuttgart, 1974).
6. Weisskopf, V. F. *Science* **187**, 605 (1975).
7. Dyson, F. J. in *Aspects of Quantum Theory* (eds Salam, A. & Wigner, E. P.) 213 (Cambridge University Press, 1972).
8. Chandrasekhar, S. *Mon. Not. R. astr. Soc.* **95**, 207 (1935).
9. Press, W. H. preprint (1973).
10. Salpeter, E. E. *Astrophys. J.* **140**, 796 (1964).
11. Rees, M. J. *Mon. Not. R. astr. Soc.* **176**, 483 (1976).
12. Silk, J. I. *Nature* **265**, 710 (1977).
13. Sciama, D. W. *Mon. Not. R. astr. Soc.* **113**, 34 (1953).
14. Rees, M. J. & Ostriker, J. P. *Mon. Not. R. astr. Soc.* **179**, 541 (1977).
15. Silk, J. I. *Astrophys. J.* **211**, 638 (1977).
16. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341 (1978).
17. Rees, M. J. *Nature* **275**, 35 (1978).

18. Carr, B. J. *Mon. Not. R. astr. Soc.* **181**, 293 (1977).
19. Zeldovich, Y. B. *J.E.T.P. Lett.* **24**, 25 (1977).
20. Liang, E. preprint (1977).
21. Zeldovich, Y. B. *Sov. Phys. Uspekhi* **20**, 945 (1978).
22. Isham, C. J., Salam, A. & Strathdee, J. *Phys. Rev.* **D3**, 1805 (1971).
23. DeWitt, B. S. *Phys. Rev. Lett.* **13**, 114 (1964).
24. Carter, B. *Large Numbers in Astrophysics and Cosmology* Paper presented at Clifford Centennial Meet., Princeton (1970).
25. Salpeter, E. E. *Astrophys. J.* **115**, 326 (1952); *Phys. Rev.* **107**, 516 (1967).
26. Hoyle, F. *Astrophys. J. Suppl.* **1**, 121 (1954).
27. Kahn, F. D. in *Emerging Universe* (eds Jacobs, K. C. & Saslaw, W. C.) 71 (Virginia University Press, 1972).
28. Weinberg, S. *Phys. Rev. Lett.* **19**, 1264 (1967); Salam, A. in *Elementary Particle Theory* (ed. Svartholm, N.) (Almqvist and Wiksell, Stockholm, 1968).
29. Bruenn, S. W., Arnett, W. D. & Schamm, D. N. *Astrophys. J.* **213**, 213 (1977).
30. Everett, H. *Rev. mod. Phys.* **29**, 454 (1957).
31. Wheeler, J. A. in *Gravitation* Ch. 44 (1971).
32. Patton, G. M. & Wheeler, J. A. in *Quantum Gravity* (eds Isham, C. J., Penrose, R. & Sciama, D. W.) Ch. 9 (Oxford University Press, 1975).

articles

The possibility of superheavy elements in iron meteorites

Leona Marshall Libby

Environmental Science and Engineering Department,

W. F. Libby

Department of Chemistry, Environmental Science and Engineering Department, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024

S. Keith Runcorn

Institute of Lunar and Planetary Sciences, School of Physics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK

The possible role of superheavy elements in the early Moon is reviewed. The siderophilic superheavy elements (possibly $Z = 114, 115, 116$) may have been the heat source in the Moon's iron core needed to generate the ancient lunar magnetic field. It is suggested that in the search for present day evidence of the existence of these hypothetical elements, iron meteorites may be a key. The available data on their trace elements are reviewed and future work to determine whether they are in part fission products from metallic superheavy elements is suggested.

WE proposed recently¹ that superheavy elements (SHE) were present in the Moon at its accretional origin, 4.6×10^9 yr ago, in sufficient abundance to melt it in its entirety—thus forming a small iron core as well as the anorthositic highlands. Geochemists at first thought of a differentiation of the outer Moon only², which it was plausible to suppose might have been brought about by accretion³ or solar wind heating⁴, processes which are most effective in the outer layers. Lunar palaeomagnetism, interpreted as being acquired in a lunar field of internal origin at the time of the cooling of the lunar rocks, $3.2\text{--}4.0 \times 10^9$ yr ago, requires the existence of a molten iron core during this time⁵. Palaeointensity experiments show the surface lunar field to have been about 1 G 4.0×10^9 yr ago, a surprisingly high value, especially as the lunar core cannot be more than 500 km in radius. Fundamental thermodynamics of the convecting core dynamo show that a heat source of 10^{12} W is required in the iron core at that time to generate the inferred magnetic field⁶. It is shown that the various conventional heat sources, proposed over the last 30 yr for the geodynamo and shown to be quantitatively sufficient for its working, are not powerful enough for the early Moon.

Indeed we¹ suggested that those superheavy elements which were siderophilic ($Z = 107\text{--}111, 114, 115, 116$) might provide the necessary heat source for the ancient lunar core dynamo. Supposing that the SHE completely melted the Moon, they could account for the differentiation, which was completed by 4.4×10^9 yr, forming the two outer lunar shells of the outer anorthositic highland layer and the source region of the mare basalts, and they could have then been incorporated into the iron core.

In Fig. 1, dN/dt is the number of SHE fissions per year per kilogram of the Moon. It was shown that 3×10^8 nuclei per year per kg are required⁷ to fission at 4.6×10^9 yr ago to melt the

Moon, and 1.1×10^7 , 6.9×10^5 and 1.6×10^4 fissions per yr per kg are required to run the dynamo at 4.0×10^9 yr, 3.6 and 3.2×10^9 yr respectively. As these values of dN/dt lie on an exponential curve of 10^8 yr half life, SHE of this half life would explain the decay of heat sources in the early Moon.

Figure 1 also shows the rate of fission of SHE nuclei in the Moon's core, assumed to be 500 km in radius, expressed per kilogram of iron for the period 4.0×10^9 yr– 3.2×10^9 yr for which the palaeointensity of the field is known. Assuming the half life of 10^8 yr previously obtained from these data, another ordinate scale can be drawn giving the abundances of SHE in the iron at various times. If the palaeointensity of the younger rocks is underestimated, as has been suggested, then it is possible to produce a plot in which the present day abundance of SHE is of the order of that claimed by Flerov⁸ in carbonaceous chondritic meteorites. The half life given by this procedure is 1.6×10^8 yr and an ordinate scale giving abundances can be drawn corresponding to this value. Also it was shown that this empirical relation gave as the original abundance of SHE in the Moon about 0.1 that of cosmic abundance of uranium.

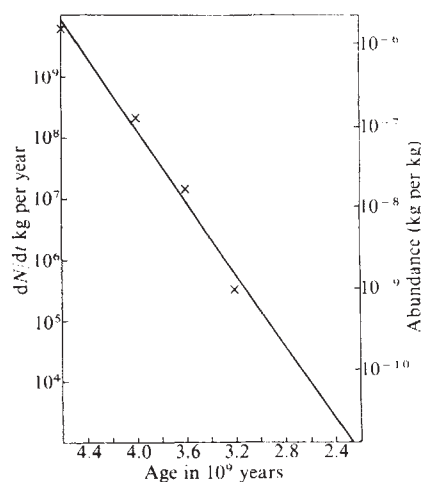


Fig. 1 Rate of fission of superheavy elements versus age in the early Moon (ordinate scale, left-hand side). Abundance by weight of the superheavy elements versus age in the Moon (ordinate scale, right-hand side).

Trace elements in iron meteorites

We have reviewed¹ the many efforts that have been made to search for evidence of SHE in nature. In this article we direct attention to the iron meteorites. We think that it is remarkable that bodies of all sizes in the Solar System (parent bodies of meteorites, the asteroid Vesta, the terrestrial planets, and the Moon) show evidence of early melting and it is a plausible hypothesis that the cause was the same. When early melting of orbiting bodies in the Solar System occurred, siderophilic elements would have accumulated in the iron phase. Half lives of 10^8 yr leave little chance of detecting SHE in iron meteorites today, but the fission products should still remain in the iron.

Figure 2 shows measurements⁹ made in iron meteorites of concentrations of elements heavier than iron. Peaks at $Z \sim 43$, $Z \sim 52$ and $Z \sim 78$ with maximum concentrations of about 10 p.p.m. (kg of element/kg of iron) are clearly present: we consider these peaks to be significant, despite the paucity of measurements and the difficulties in the mechanical separation of inclusions, such as sulphides and silicates, from the iron phase. Figure 3 shows that there is evidence for two of the peaks in the more primitive carbonaceous chondrites; they also occur (with much smaller amplitude relative to iron) in the Cameron¹⁰ curve of abundance of elements in the primitive solar nebula around $Z = 53$ and 78 (magic numbers) where the peaks have an abundance roughly 10^{-6} relative to iron. We must therefore conclude that the iron meteorites have not taken up, either in the process of condensation from the solar nebula or from any late processes of differentiation and melting in the 'parent body', any appreciable amounts of the other elements. This has been explained following Goldsmidt¹¹ by classifying the former elements as siderophile and the latter as lithophile; we attempt here a quantitative treatment. To establish whether the shapes of these peaks are entirely explained by the solubilities in iron of the elements originally present in the solar nebula we have made

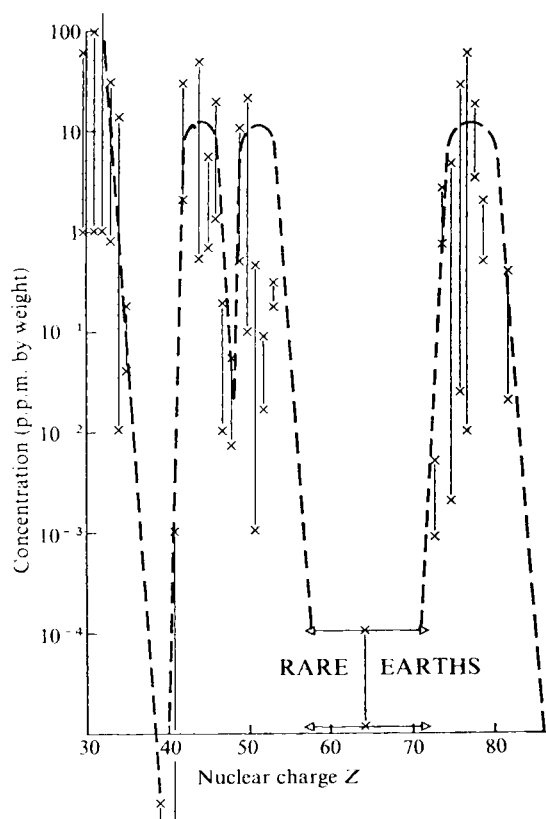


Fig. 2 Abundance of elements measured in iron meteorites versus Z (see ref. 9). Dashed line shows suggested fission product distribution.

the following computation. We assume that the liquid metals obey the perfect solution law, that is, the solubility S_s is equal to the reciprocal of the ratio of the vapour pressure of the pure solute to the ambient pressure over the solution. Based on this assumption we have computed the expected concentration of the metallic elements in iron at the freezing point of iron.

We further assume that if an element's oxide is stable at $1,535^\circ\text{C}$, the melting point of iron, it will not be miscible with iron and so its concentration in iron meteorites will be zero. This assumption produces zeros in the theoretical curve of concentration against Z , from $Z = 33$ to $Z = 39$ inclusive and from $Z = 52$ to 71 inclusive. For the light elements this assumption is also true, and, in particular aluminium oxide is very stable and will never be found in iron: thus ^{26}Al on our theory is not plausible as a heat source in an iron core as has been suggested¹². (In any case its short half life of 0.73×10^6 yr rules it out as a heat source for the Moon's dynamo.) Similarly, halogens form hydrogen halides in a hydrogen-rich solar nebula and are immiscible in iron therefore. We have further assumed that the rare gases are insoluble in iron.

For those elements which are not rare gases, not halogens, and which do not have very stable oxides, their heats of vaporisation ΔH are computed using Trouton's rule: $\Delta H = (25) \times (\text{boiling point, K}) \text{ cal mol}^{-1}$. Putting this value of ΔH into the Clausius-Clapeyron equation, P_s , the vapour pressure of the given element as pure liquid at the freezing point temperature of iron (chosen as the lowest temperature at which a solute element could dissolve in iron) is calculated. Similarly the vapour pressure P_{Fe} of iron at its melting point temperature is computed. Assume a solar nebula at 1 atm pressure, containing 10^{12} atoms H, $10^{6.9}$ atoms Fe (designated as $n_{\text{Fe-O}}$), and numbers of atoms of each element (designated as n_{s-O}) equal to its cosmic abundance relative to hydrogen at 10^{12} . Let the number of solute atoms in the vapour at the melting point of iron be $n_s = P_s V/kT$ and the number of iron atoms in the vapour similarly be $n_{\text{Fe}} = P_{\text{Fe}} V/kT$. If S_s is the mole fraction of solute in the condensed system at the freezing point of iron, $S_s = P_s/P_{s-O}$. Then, using conservation of numbers of atoms,

$$n_{s-O} = n_s + (n_{\text{Fe-O}} - n_{\text{Fe}}) \times S_s / (1 - S_s)$$

which, with some rearrangement, yields the relation (using $V/kT = 10^{12}$ and using $S_s \ll 1$),

$$S_s = n_{s-O} (10^{12} + (n_{\text{Fe-O}} - n_{\text{Fe}}) / P_{s-O})^{-1} \times (P_{s-O})^{-1}$$

where $(N_{\text{Fe-O}}) = (n_{\text{Fe-O}})$ is the number of condensed iron atoms.

Simple reasoning by analogy leads one to guess that the SHE elements of lower Z (114, 115, 116) would be siderophilic, whereas elements of $Z = 121$ to 138 probably would have a chemistry like that of actinium and lanthanum, and that they would form stable oxides in a water-rich nebula and could not have existed in iron meteorites. However, elements of $Z = 121$ –138 are g -orbital elements of which we have no chemical experience. So it is possible that our capability of predicting their solubility in iron fails in this range of Z . Elements 107–111 almost certainly will be iron soluble.

Table 1 lists the theoretical solubilities in iron versus Z , calculated by the method discussed above and the measured concentrations versus Z . In general the calculated solubilities in iron fit the observations in iron meteorites for about half the metallic elements, which have long been recognised as siderophiles, but they are lower than the observed ones in most other cases. The differences in concentrations—those found experimentally minus those calculated—provide evidence of strong peaks around $Z = 32$, $Z = 46$, $Z = 50$ and $Z = 77$. These differences, plotted in Fig. 4, agree well with the peaks in the experimental measurements of heavy elements in iron meteorites. Furthermore, the highest values of peak elements occur in those metal meteorites with the largest crystals having cooled the most slowly. In addition, we have evidence for the presence of Zn, the halogens, the noble gases, the rare earths, and the actinides, none of which should be present in an iron phase according to our theory. Experimental confirmation of the

Table 1 Computed solubility of heavy elements in iron compared with measured concentrations of heavy elements in iron meteorites (see ref. 9)

Element	Charge	Calculated kg per 10^6 kg Fe	Maximum observed kg/ 10^6 kg of Fe	Excess concentration of element*	Element	Charge	Calculated kg per 10^6 kg Fe	Maximum observed kg/ 10^6 kg of Fe	Excess concentration of element*
Zn	30	—	61	61	Ba	56	Zero	—	—
Ga	31	0.14	100	100	La	57	—	—	—
Ge	32	8.2	520	506	to	—	Zero	10^{-4} to 10^{-6} (?)	10^{-4} to 10^{-6}
As	33	Zero	30	30	Lu	71	—	—	—
Se	34	Zero	—	—	Hf	72	2.8	—	—
Br	35	Zero	0.15	0.15	Ta	73	2.8	5×10^{-3}	-2.8
Kr	36	Zero	10^{-6}	10^{-6}	W	74	5.3	2.7	-2.5
Rb	37	Zero	—	—	Re	75	1.7	4.8	3
Sr	38	Zero	—	—	Os	76	8.5	50	42
Y	39	Zero	—	—	Ir	77	6.9	59	52
Zr	40	52	—	—	Pt	78	17	19	2
Nb	41	10	<20	Zero	Au	79	0.29	2.2	2
Mo	42	18	11	-7	Hg	80	8.2×10^{-9}	2.2 (?)	2.2
Tc	43	Zero	—	—	Tl	81	6.6×10^{-6}	0.25×10^{-3}	0.0025
Ru	44	7.7	35	27	Pb	82	6.7×10^{-4}	0.4	0.4
Rh	45	1.3	4.0	3	Bi	83	2.1×10^{-9}	—	—
Pd	46	4.6	19	14	Po	84	Zero	—	—
Ag	47	10×10^{-4}	0.2	0.2	At	85	Zero	—	—
Cd	48	13×10^{-3}	0.056	0.04	Rn	86	Zero	—	—
In	49	4.1×10^{-4}	0.02 (?)	0.02 (?)	Fr	87	Zero	—	—
Sn	50	0.013	20 (?)	20	Ra	88	Zero	—	—
Sb	51	12.6	0.5	-12.1	Ac	89	Zero	—	—
Te	52	Zero	—	—	Th	90	Zero	1.8×10^{-5}	1.8×10^{-5}
I	53	Zero	0.3 (?)	0.3	Pa	91	Zero	—	—
Xe	54	Zero	10^{-6} (?)	10^{-6} (?)	U	92	Zero	10^{-4} to 10^{-6} (?)	10^{-4} to 10^{-6}
Cs (?)	55	Zero	—	—					

* Excess is the difference between the measured and calculated concentrations in p.p.m. by weight.

theoretical calculations is lacking and is much needed.

From the point of view of physical chemistry, these results are only a first approximation. The basic consideration is whether the assumption of equilibrium with the nebular gases at the time of freezing of the iron rich phase is valid. Many workers on iron meteorites assume that further fractionation via formation of 'parent bodies' and subsequent break-up to form the meteorites occurs. We suggest that this may not be so and that the slow cooling required by the Widmanstätten etch patterns may have been due to the presence of superheavy elements furnishing heat by radioactive decay, thus locking the cooling rate into their radioactive lifetimes.

Fission of superheavy elements

The sum of the nuclear charges of the two middle mass peaks equals $Z = 94-96$; but because the oxides of uranium and plutonium are very stable, these elements cannot be present in iron meteorites in quantities sufficient to explain these two peaks as the products of their fission: nor does the measured concentration of uranium in iron meteorites permit this explanation as will be discussed below.

The sum of the nuclear charges for the peaks at $Z \sim 46$ and $Z \sim 78$ is 121, and the sum for those at $Z = 52$ and $Z = 78$ is 130. Thus, if two superheavy elements had produced these peaks by fission, their nuclear charges would have ranged from 110 to 120. The width of the peak at $Z \sim 78$ is equal to twice the width of the peak at $Z \sim 43$. For this reason we have drawn the second peak, at $Z \sim 52$, in such a way that it has the same width as the peak at $Z \sim 43$, to account for the enhanced width at $Z \sim 78$, suggesting that indeed two superheavy elements are the progenitors. However, some theorists^{13,14} prefer superheavy elements in the low Z end of the island of stability, $Z = 114$, 115, 116, to be the most stable and these, as we have seen, are the only ones, on present chemical reasoning, to be siderophile.

The sum of the charges of three peaks in Fig. 2 corresponds to $Z = 168-180$. Since nuclear shell theory predicts a further island of stability at $Z = 164-184$ (depending on which potential energy function is favoured)¹⁵, we must consider whether the three peaks originate from decay of super-superheavy elements

(SSHE) by ternary fission. Although uranium and plutonium decay by ternary fission with only very low probability, it is possible that the geometry of the super-superheavy elements is such that ternary fission becomes the most probable mode of decay. The shape of the nucleus in its ground state could then be deduced from the shapes of the peaks of the fission product distribution. The shape might, for example, be bent and lumpy, perhaps like a water molecule, or perhaps like a three-legged starfish. According to recent theoretical ideas¹⁶, such nuclei

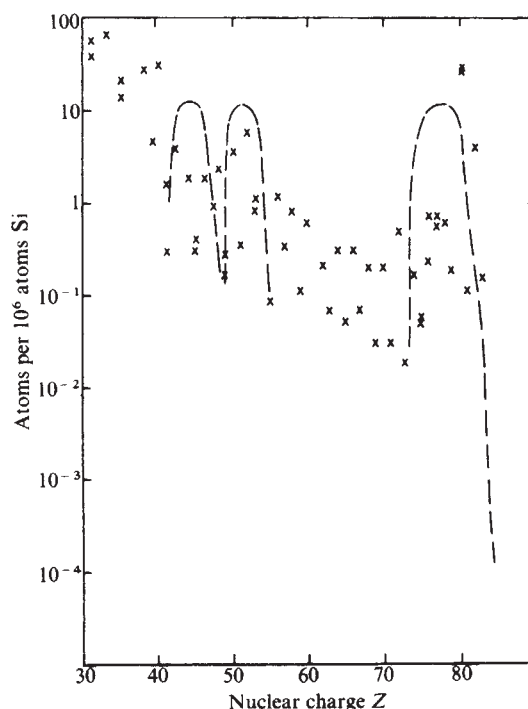


Fig. 3 Excess abundance of elements in iron meteorites above that expected on solubility theory.

($Z \sim 173$, $N \sim 249$) could have three axes of rotational symmetry.

Our hypothesis that the peaks result from the decay of SHE and SSHE in the early Solar System encounters some quantitative difficulties. If energies released on fission by the SHE and SSHE are ~ 250 MeV, then the peaks of concentration at about 10^{-5} g g $^{-1}$ would imply that in the early cores of the parent bodies a total energy release of about 2×10^4 J kg $^{-1}$, sufficient to vaporise iron. However, radiative cooling would prevent this. At the melting point of iron, $1,535^\circ\text{C}$, and with a mean life of Y yr for spontaneous fission, an iron sphere taken as a black body would radiate at the fission rate for any values of Y longer than 1 Myr for spheres of radii less than 3,000 km. However, from the studies of the Widmanstätten patterns and kamacite bands¹⁷, we know that the cooling rates of iron meteorites must have been of the order of 1°C Myr^{-1} . This probably requires the parent bodies of the iron meteorites to have been a loosely held aggregate in order to allow radiative cooling. Such a model for them may be required to explain why they were broken up by external forces, possibly tidal. We conclude that for iron meteorites early solidification is probable, with sealing-in of a substantial part of the fission products. Scott and Wasson¹⁷ discuss the classes of iron meteorites: only some may have originated as cores of small planetary bodies. The data we have used include all iron meteorites together; plots like that of Fig. 2 should be made for each class when sufficient measurements are available.

Another difficulty for the hypothesis is that the peak concentrations of the hypothetical fission products are about 10^{-5} kg kg $^{-1}$ and would imply, assuming a maximum fission yield for an element of 5%, that the initial abundance of SHE in the early Solar System, allowing that they became concentrated in iron by a factor of 10–100 would have been between 2×10^{-6} and 2×10^{-5} kg kg $^{-1}$. This is to be compared with a cosmic abundance of 10^{-7} for uranium. Recent calculations of the r process suggest a cosmic abundance of SHE within an order of magnitude that of uranium¹⁸. If this were so, and if the SHE most likely to have been present were those with $Z = 114$ for the two reasons already given, then the fission peaks expected by analogy with uranium should be near $Z = 50$ and $Z = 80$ and with abundance 5×10^{-8} – 5×10^{-7} . If the peaks which we have

discussed are after all the result of elements being taken up from the solar nebula by fractionation processes and as a result of extraction of elements into the iron phase during iron-silicate fractionation, then, as there is a window around $Z = 55$ – 70 where the expected fission peaks should lie, better data on rare earth abundances should enable their presence to be determined. So far virtually no data are available.

In iron meteorites the abundance of uranium is 10^{-10} – 10^{-12} (see ref. 19), completely inadequate to explain the 10 p.p.m. of lead and the several parts per million of Sn, Sb, Te and I found in iron meteorites and cannot explain the Xe and Kr which has recently been detected^{20,21} in iron meteorites with concentrations of about 10^{-12} because their isotopic ratios are not those of fission products.

In the so-far unsuccessful attempts to make SHE in accelerators, collisions of heavy ions have been used²². We note that in collisions of ^{40}Ar (~ 300 MeV) and ^{84}Kr (~ 600 MeV) with ^{238}U , three peaks of heavy element concentrations were produced²² at about the same values of Z as those of the peaks for the iron meteorites (Fig. 2). It is as if perhaps SHE of $Z = 110$ – 120 were being produced but their neutron/proton ratio was too small and their excitations too large to allow them to live long after the collision.

If indeed the peaks are made by fission products of SHE in iron meteorites, the wide separations of the peaks indicates a very asymmetric mode of binary fission. Asymmetric fission was predicted by Rao²³ at the very values of Z and N which characterise the peaks in Fig. 2. They centre around the 'magic numbers': for the peak at $Z = 46$, the magic number is $N^* = 50$; the peak at $Z = 50$ centres around the two magic numbers, $N^* = 82$ and $Z^* = 50$; and the peak at $Z = 78$ centres around the two magic numbers $N^* = 126$ and $Z^* = 82$. We believe that decay of superheavy elements initially present in supernova remnants is partly responsible for the peaks mentioned above in the carbonaceous chondrite abundances. It would be strange if the abundance peaks observed in the iron meteorites were to be entirely explained by their siderophile character. Why would elements with magic numbers in their nuclear structure also have the chemical property of solubility in iron?

Classical theory of iron meteorites

E. R. D. Scott²⁵ states that "all major groups (of iron meteorites) were produced by fractional crystallisation of an iron melt. Meteorites did not form a single molten core. The cooling rates estimated from kamacite bandwidths and bulk nickel contents do not support a core origin from some groups. The wide range of cooling rates suggested formation of these meteorites in isolated regions of a parent body. Metre-sized molten iron pools formed. Unless it can be shown that other factors (such as the effect of trace and minor elements on diffusion rates) are responsible for the wide range of cooling rates estimated, such data are not compatible with core formation. The primary fractionation which established the chemical differences between the groups might have occurred during condensation of the solar nebula. It seems unlikely that the major differences could have been produced during separation of metal, silicate, and sulphide phases in parent bodies."

J. T. Wasson²⁶ suggests that "All currently available evidence favours the viewpoint that meteorites formed at relatively low ambient pressures, much less than those found in the interior of a moon-sized body, and consistent with an origin in parent bodies no larger than the largest asteroids. Now that a low-pressure origin of the meteorites is widely accepted, ... some, perhaps many, probably originated in the smaller asteroids, with central pressures less than 100 bar ... the minimum size of the parent body can be estimated, the total mass is 0.8×10^{15} g, 650 Myr ago. At a density of 7.9 g cm $^{-3}$, the minimum radius is 0.3 km. This lower limit is much smaller than parent body sizes necessary to account for estimated cooling rates." Thus while we acknowledge that many people have worked intensively to understand the mysterious properties of iron meteorites for at

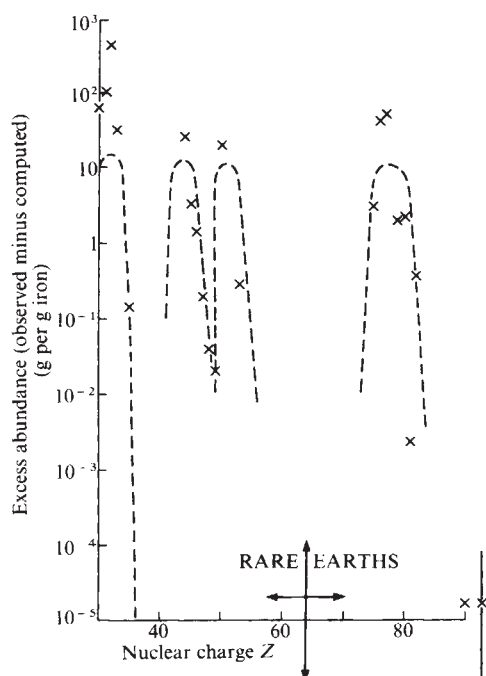


Fig. 4 Abundance of elements in carbonaceous chondrites. Dashed line shows suggested fission product distribution.

least 40 yr, many of the properties of iron meteorites are inexplicable by the classical ideas of their origin. We hypothesise that the iron meteorites were never very large, condensed from a solar nebula in small masses containing superheavy elements soluble in iron-nickel melts, cooled very slowly because of a radio-half life of about 100 Myr, and produced solid iron-nickel chunks supersaturated in fission products which ranged from $Z = 30$ to 92, and perhaps an even wider distribution of elements of charge Z .

Conclusions

Our hypothesis that the heavy elements in iron meteorites are partly the fission products of SHE and perhaps of SSHE is put forward tentatively as befits our state of ignorance in this field. But we suggest some definitive measurements to test it: (1) measurements of heavy element concentrations in each class of iron meteorites and of halides, alkalis, alkaline earths, and rare gases in particular; (2) measurement of solubilities of heavy elements in iron versus pressure and at high temperatures, to test our crude calculations; (3) study of the theory of ternary fission in SSH elements; (4) bombardment of uranium with uranium ions²³ in the heavy ion accelerators to determine if the heavy element peak at $Z \sim 78$ increases in height relative to that observed in Ar-U and Kr-U bombardments²⁴, and bombardment of U with Xe ions, for the same purpose; and (5) further work on isotope ratio analyses of peak elements.

We thank R. Hutchison, J. Kerridge, E. R. D. Scott and J. Wasson for discussions.

Received 18 April 1978; accepted 29 January 1979.

1. Runcorn, S. K., Libby, L. M. & Libby, W. F. *Nature* **270**, 676-681 (1977).
2. Taylor, S. R. *Lunar Science: A Post Apollo View* (Pergamon, Oxford, 1975).
3. Ransford, C. A. *Lunar Sci. Abstr.* **VIII**, 793-794 (1977).
4. Sonett, C. P., Colbourne, D. S. & Schwartz, K. *Icarus* **24**, 231-255 (1975).
5. Collinson, D. W., Stephenson, A. & Runcorn, S. K. *Phil. Trans. R. Soc. A* **285**, 241-247 (1977).
6. Runcorn, S. K. *Science* **199**, 771-773 (1978).
7. Runcorn, S. K. *Earth planet. Sci. Lett.* **39**, 193-198 (1978).
8. Flerov, G. N., Ter Akopjan, G. M., Popecko, A. G., Fefilov, B. V. & Subbotin, V. G. *Sov. J. Nuclear Phys.* **26**, 449 (1977).
9. *Handbook of Elemental Abundances in Meteorites* (ed. Mason, B.) (Gordon and Breach, New York, 1971).
10. Cameron, A. G. W. *Space Sci. Rev.* **15**, 121-146 (1973).
11. Goldsmid, V. M. *Geochemistry* (Oxford University Press, 1939).
12. Singer, C. E. *Nature* **272**, 239 (1978).
13. Nilsson, S. G. *et al. Nuclear Phys. A* **131**, 1-66 (1969).
14. Fiset, E. O. & Nix, J. R. *Nuclear Phys. A* **193**, 647-671 (1972).
15. Sobicewski, A. *Nobel Symp.* **27**, 47-52 (1974).
16. Bohr, A. & Mottelson, B. R. *Nobel Symp.* **27**, 13-20 (1974).
17. Scott, E. R. D. & Wasson, J. *Rev. Geophys. Space Phys.* **13**, 527-546 (1975).
18. Schramm, D. N. & Fowler, W. A. *Nature* **231**, 103-106 (1971).
19. Reed, G. W., Hamaguchi, H. & Turkevich, A. *Geochim. cosmochim. Acta* **13**, 248-255 (1958).
20. Munk, M. M. *Earth planet. Sci. Lett.* **2**, 301-309 (1967).
21. Munk, M. M. *Earth planet. Sci. Lett.* **3**, 133-138 (1967).
22. Seaborg, G. T. *Nobel Symp.* **27**, 5-12 (1974).
23. Rao, M. N. *Nuclear Phys. A* **140**, 69-73 (1970).
24. Seaborg, G. T. *The Transuranium Elements, Proceedings of the Robert A. Welch Foundation*, 5-40 (1969).
25. Scott, E. R. D. *Geochim. cosmochim. Acta* **36**, 1205-1236 (1972).
26. Wasson, J. T. *Meteorites* (Springer, Berlin, 1974).

Transfer of actinides from the English Channel into the southern North Sea

C. N. Murray*, H. Kautsky & H. F. Eicke

Deutsches Hydrographisches Institut, 2000 Hamburg 4, FRG

Results are presented for ^{238}Pu , $^{239+240}\text{Pu}$, ^{241}Pu , ^{241}Am , ^{242}Cm and ^{244}Cm in waters of the English Channel and southern North Sea. Based on actinide concentration ratios in these waters compared with expected fallout values, the activities measured are considered to be mainly due to releases of low-level waste from the La Hague fuel reprocessing plant. Measurements made on ^{242}Cm indicate that any support by $^{242\text{m}}\text{Am}$ would be minimal.

AS part of a study of the distribution, behaviour and sources of actinide isotopes in the North Sea, two series of surface water samples were taken during February-March and September-October 1975, concentrating especially on an area of the southern North Sea and the English Channel. This area is of particular interest because of the situation of a nuclear fuel reprocessing plant at La Hague near Cherbourg, as well as various nuclear power installations on coastal or estuarine sites in countries bordering this region. Previous work¹⁻³ has shown that important amounts of various fission radionuclides originating from releases of low-level wastes find their way into waters of the North Sea.

Murray and Kautsky⁴ have shown that waters entering the northern North Sea by the Pentland Firth and to the north of the Orkney Islands also transport detectable quantities of $^{239+240}\text{Pu}$ and ^{238}Pu resulting from discharge from fuel reprocessing plants at Windscale and to a much lesser extent, Dounreay. We report here the results of measurements of isotopes of three actinide elements, plutonium, americium and curium, in waters of the English Channel, whose activities are also believed to result

from waste discharge. These waters act as input source of several isotopes into the southern North Sea. Comparisons of various actinide activity concentration ratios for these waters with those that may be expected from fallout have been used to identify the source and the extent of their penetration.

Sampling and analysis

Samples of surface seawater, varying in volume between 60 and 200 l, were taken at 13 stations in the Channel and North Sea between 50° and 55° N during February-March, and September-October 1975. The samples were immediately acidified on collection and stored unfiltered in plastic drums until their subsequent analysis. Separation and measurement of the samples were carried out within three months of collection. The chemical separation and measurement by α spectrometry of the various actinide isotopes and the measurement of IAEA intercalibration samples have been described elsewhere⁵. After initial measurements of $^{239+240}\text{Pu}$, ^{238}Pu and ^{241}Am , the disks containing plutonium were stored for nearly two years so that ^{241}Pu (β emitter) could be measured from the ingrowth of its daughter isotope ^{241}Am (for a detailed discussion of this procedure see ref. 6). Due to the short time between sample collection and analysis, initially separated ^{241}Am is assumed to contain a negligible contribution of ^{241}Am due to ^{241}Pu decay between collection and analysis. Where activity levels were sufficiently high, measurements were also made of ^{242}Cm and ^{244}Cm . The decay of ^{242}Cm was traced on a single sample for nearly three years to determine whether some fraction of it was supported by $^{242\text{m}}\text{Am}$.

Ratio of actinides in fallout

Hardy *et al.*⁷ have discussed the global inventory and distribution of fallout plutonium caused by atmospheric nuclear-weapon testing; from their data an average $^{238}\text{Pu}/^{239+240}\text{Pu}$ ratio

* Present address: Commission of the European Communities, JRC Ispra Establishment, Chemistry Division, 21020 Ispra (Varese), Italy.

Table 1 Seawater activities of actinides (f kg^{-1}) in surface waters of the English Channel and southern North Sea

Cruise	Local station no.	^{238}Pu	$^{239+240}\text{Pu}$	^{241}Pu	$^{241}\text{Am}^\dagger$	$^{242}\text{Cm}^\ddagger$	^{244}Cm
Feb–March 1975	52	2.0 ± 0.1	10.6 ± 0.3	364 ± 31	1.48 ± 0.09	7.48 ± 0.7	ND
	53	1.4 ± 0.1	10.5 ± 0.4	493 ± 42	NM	NM	NM
	56	1.5 ± 0.2	8.4 ± 0.7	78 ± 17	0.28 ± 0.04	2.9 ± 0.3	ND
	60	ND	2.3 ± 0.4	NM	0.20 ± 0.04	ND	ND
	62	0.2 ± 0.05	1.9 ± 0.2	53 ± 15	0.10 ± 0.03	ND	ND
	63	0.3 ± 0.1	1.6 ± 0.2	39 ± 17	0.10 ± 0.03	ND	ND
	73	$0.03 \pm 0.01^*$	0.4 ± 0.1	11 ± 9	0.04 ± 0.02	ND	ND
	1	$0.06 \pm 0.02^*$	0.8 ± 0.1	14 ± 12	0.04 ± 0.01	ND	ND
Sept–Oct 1975	30	0.98 ± 0.19	3.7 ± 0.3	113 ± 25	0.76 ± 0.08	1.68 ± 0.23	0.07 ± 0.02
	31	0.65 ± 0.10	2.8 ± 0.3	140 ± 20	0.36 ± 0.05	0.39 ± 0.07	0.03 ± 0.01
	34	0.35 ± 0.04	2.5 ± 0.1	100 ± 15	0.32 ± 0.04	0.46 ± 0.02	ND
	22	0.33 ± 0.05	1.5 ± 0.1	88 ± 13	0.13 ± 0.03	ND	ND
	40	0.26 ± 0.05	1.8 ± 0.1	75 ± 14	0.23 ± 0.06	ND	ND

Error, 1σ propagated; NM, not measured; ND, not detected.

*Estimated activity for coastal water (ref. 17).

†No correction for ingrowth of ^{241}Am from ^{241}Pu as sample collection date to analysis was ≤ 3 months only.

‡Corrected for decay to date of collection.

of 0.04 can be calculated for integrated soil samples for 1970–71. They estimate that the accidental release of ^{238}Pu from the SNAP 9A satellite almost tripled the global deposition of this isotope by 1970.

Livingston *et al.*⁶ have reported that the $^{241}\text{Pu}/^{239+240}\text{Pu}$ ratios for fallout-contaminated marine samples are similar to those from fallout-contaminated soils. They showed, however, that these ratios are clearly different in samples from planned or accidental plutonium releases. Whereas the global fallout activity $^{241}\text{Pu}/^{239+240}\text{Pu}$ ratio on 1 January 1970 was calculated to be 8.2, the ratio in IAEA intercalibration samples, which are known to originate from areas contaminated by waste effluent from a major fuel reprocessing plant, was found to be about 30.

Fukai *et al.*⁸ measured the activity ratio of $^{241}\text{Am}/^{239+240}\text{Pu}$ in Mediterranean surface waters sampled in 1975 as 0.055 ± 0.007 . Results¹⁰ for North Atlantic surface water gave a range of ratios of 0.05–0.28, which agree with the estimate of Krey *et al.*¹¹ for 1974 for integrated global fallout of 0.22. Livingston *et al.*⁶ calculated the increase of the $^{241}\text{Am}/^{239+240}\text{Pu}$ ratio due the decay of fallout ^{241}Pu with time and for 1975 a value of about 0.18 was determined.

Holm and Persson⁹ and Bowen and Livingston¹² have discussed the occurrence of ^{242}Cm and ^{244}Cm in the environment. The former report that the average $^{242}\text{Cm}/^{239+240}\text{Pu}$ ratio in lichen for the period 1961–72 was $(0.3 \pm 0.1) \times 10^{-4}$. They also estimated the $^{244}\text{Cm}/^{239+240}\text{Pu}$ activity ratio to be about 0.2×10^{-4} , which agrees well with the observations of Bowen and Livingston¹². Both groups of workers assumed that ^{242}Cm was supported by ^{242m}Am , due to the age of the samples and the relatively short half life of ^{242}Cm (163 d).

On the basis of these expected fallout ratios an attempt has been made to distinguish the probable source and dispersion of actinides entering the southern North Sea from the English Channel.

Results and discussion

Figure 1 shows the positions of the sampling stations measured during February–March, and September–October 1975. Results for surface water activities of ^{238}Pu , $^{239+240}\text{Pu}$, ^{241}Pu , ^{241}Am , ^{242}Cm and ^{244}Cm are presented for both cruises in Table 1; all activities are corrected to the date of sampling. The concentrations for ^{238}Pu , $^{239+240}\text{Pu}$ and ^{241}Am for the February–March cruise have been presented elsewhere⁴ but these results are needed for calculation of isotopic ratios and are, therefore, included.

Kautsky¹³ has shown that patches of water of greatly varying ^{137}Cs activity enter the southern North Sea from the English

Channel at irregular intervals, rather than as a steady input. He proposed that this probably reflected irregular discharge by the La Hague fuel reprocessing plant. Such irregular discharges probably characterise the input of actinides as well.

As there are no available data for the variations in amounts of radionuclides discharged by the French plant as a function of time, a difficulty arises in trying to separate differential actinide behaviour due to chemical reactivity (towards biological or sedimentological processes) from discharge compositional fluctuations and contributions from fallout.

An indication of the pulse-like nature of the actinide discharge in the English Channel may be obtained by comparing

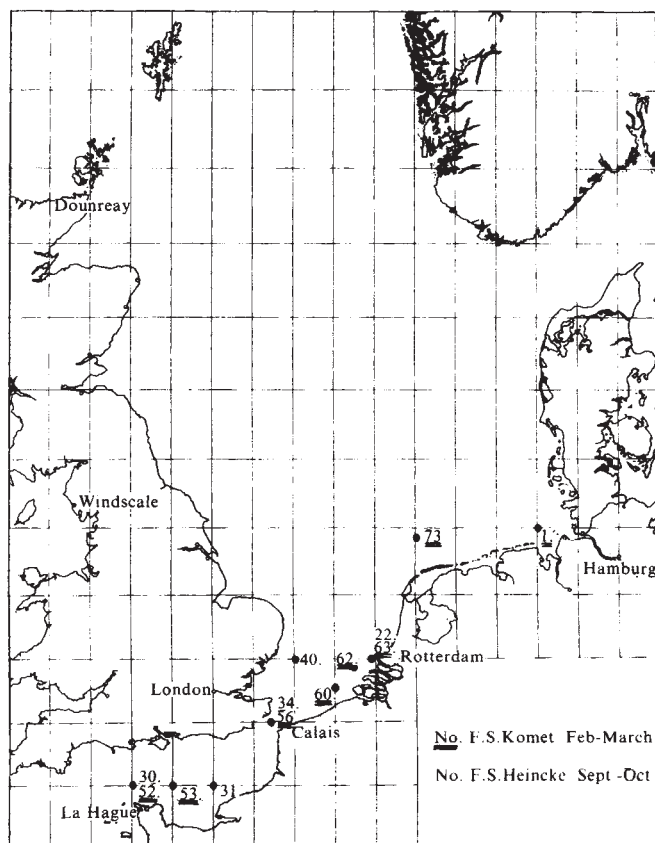


Fig. 1 Sampling stations February–March, and September–October 1975 cruises.

Table 2 Variations in channel water activity ratios occurring between February–March, and September–October 1975 cruises

Stations	^{238}Pu	$^{239+240}\text{Pu}$	^{241}Pu	^{241}Am	^{242}Cm
March–October					
52/30	2.0 ± 0.4	2.9 ± 0.3	3.2 ± 0.8	2.0 ± 0.2	4.5 ± 0.7
53/31	2.2 ± 0.4	3.8 ± 0.4	3.5 ± 0.6	–	–
56/34	4.3 ± 0.7	3.4 ± 0.3	0.8 ± 0.2	0.9 ± 0.1	6.3 ± 0.7
Average decrease factor	2.8 ± 0.5	3.4 ± 0.3	2.5 ± 0.5	1.5 ± 0.2	5.5 ± 0.7

Error, 1σ propagated.

concentrations at stations 52–56 (February–March 1975) and 30–34 (September–October 1975): these results are shown in Table 2. Although dilution occurs during the movement of water through the Channel, the variation in activity for ^{238}Pu , $^{239+240}\text{Pu}$ and ^{241}Pu , calculated as isotope activity in March divided by isotope activity in October (assuming that the dilution will be similar for both cruises), shows a remarkably similar pattern; the activity decrease for the isotopes being 2.8, 3.4 and 2.5 respectively. Thus the plutonium isotopes' water activity in the English Channel for March 1975 seems to have been some three times greater than that occurring for October 1975.

In the case of ^{241}Am a change in activity of about 2 occurs in the region of Cap de La Hague during the same period. The activity ratio in the Straits of Dover (stations 56/34, March–October), however, is about 0.9. As the ingrowth of ^{241}Am from ^{241}Pu will affect both concentrations equally, the difference may be considered real.

The results for ^{242}Cm are more complicated because of its relatively short half life of 163 d. Kautsky¹³ estimated the average speed taken for water to pass from Cherbourg to the Straits of Dover as about 1.7 nautical miles per day. The time it would take from Cherbourg to the Straits of Dover, a distance of ~160 nautical miles, would thus take some 100 d or somewhat less than one half life of ^{242}Cm . Assuming that both dilution and decay will affect the samples from the two cruises equally, it can be seen that ^{242}Cm activity in the English Channel in March 1975 was some 5.5 times greater than in October 1975.

It thus seems that the average concentration fluctuations for plutonium isotopes in the English Channel during March–October 1975, were reasonably constant. However, when one compares the fluctuations between the elements, large variations can be seen to occur; with decrease in the order $\text{Cm} > \text{Pu} > \text{Am}$. That the activities measured in the English Channel are not primarily due to fallout but may be attributed to the discharge of low-level waste from La Hague is shown by comparing the various isotope activity ratios in Table 3 with those that can be expected from fallout. In the case of plutonium isotopes the fallout $^{238}\text{Pu}/^{239+240}\text{Pu}$ ratio for integrated samples is about

0.04 ± 0.01 (2σ error) based on the results of Hardy *et al.*⁷. The ratios found in the channel region are about 0.20 for both cases. Even in the southern North Sea as far east as about $53^\circ 00' \text{N}$, $04^\circ 00' \text{E}$, plutonium activities are clearly due to sources other than fallout.

The results for the activities of ^{241}Pu found during both periods in 1975 are considerably higher than those of $^{239+240}\text{Pu}$. The values near Cap de La Hague are 30–50 times greater and are significantly different from those expected from fallout. The ^{241}Pu activity for station 56, February–March 1975 is 3–4 times lower than for other stations in the region. The reason for this anomalously low result is not clear, however, the $^{241}\text{Pu}/^{239+240}\text{Pu}$ ratios are generally comparable with those reported for 1972–74 by Hetherington *et al.*¹⁴ who also reported a large variability in the Windscale effluent ratio of 38.

The absolute activities of ^{241}Am in Channel water for both cruises are 5–10 times greater than those for Mediterranean surface water contaminated only by fallout. Average activities of about 0.06 fCi l^{-1} have been reported by Fukai *et al.*⁸ and Livingston *et al.*¹⁵. East of the Straits of Dover it becomes difficult to characterise the source of americium in surface seawater although it is clear from the $^{238}\text{Pu}/^{239+240}\text{Pu}$ ratio that plutonium entering the southern North Sea can be seen at least as far as $52^\circ 00' \text{N}$. The significantly higher ^{241}Am activities at stations in the Channel can, however, be attributed to discharges by the French plant.

An indication of a difference in behaviour of americium and plutonium may be deduced if one considers the $^{241}\text{Am}/^{239+240}\text{Pu}$ ratio. If on the basis of Kautsky's¹³ estimated transport velocity of about 1 nm d^{-1} for water moving from Cherbourg to the German Bight, a period of about 0.7 yr is required for activity to reach $52^\circ 00' \text{N}$. The ingrowth of ^{241}Am from ^{241}Pu for a $^{241}\text{Pu}/^{239+240}\text{Pu}$ isotope ratio of 40 would produce, in this time, an *in situ* increase in the measured $^{241}\text{Am}/^{239+240}\text{Pu}$ ratio of about 0.04, supplementing that due to direct input of ^{241}Am from La Hague. Taking expected fallout $^{241}\text{Am}/^{239+240}\text{Pu}$ ratios in surface waters to range between 0.05 and 0.28, no significant differences from fallout for either the Channel or the southern North Sea can be detected for the two cruises. However, the absolute ^{241}Am activities are clearly significantly different from fallout in the Channel, compared with those in the southern North Sea where plutonium activities due to waste discharge are to be seen. Thus the rate of disappearance of americium introduced from fuel reprocessing plants into surface waters is apparently considerably faster than for plutonium. This behaviour has also been reported¹⁶ for americium discharged into the Irish Sea. In the case of fallout for Mediterranean water Livingston *et al.*¹⁵ have also shown that a more rapid loss of ^{241}Am occurs.

Only a small number of samples from the English Channel showed detectable levels of ^{242}Cm and ^{244}Cm . On the basis of a comparison of ratios of $^{242}\text{Cm}/^{239+240}\text{Pu}$ and $^{244}\text{Cm}/^{239+240}\text{Pu}$

Table 3 Activity ratios of actinides in surface waters of the English Channel and southern North Sea

Date	Local station no.	$^{238}\text{Pu}/^{239+240}\text{Pu}$	$^{241}\text{Pu}/^{239+240}\text{Pu}$	$^{241}\text{Am}/^{239+240}\text{Pu}$	$^{242}\text{Cm}/^{239+240}\text{Pu}$	$^{244}\text{Cm}/^{239+240}\text{Pu}$
Feb–March 1975	52	0.19 ± 0.01	34.2 ± 3.1	0.14 ± 0.01	0.71 ± 0.07	–
	53	0.13 ± 0.02	47.0 ± 4.4	–	–	–
	56	0.18 ± 0.03	9.3 ± 2.2	0.03 ± 0.006	0.35 ± 0.05	–
	62	0.11 ± 0.03	27.6 ± 5.9	0.05 ± 0.02	–	–
	63	0.19 ± 0.07	24.9 ± 11.1	0.06 ± 0.02	–	–
	73	$*0.06 \pm 0.03$	26.7 ± 27.0	0.10 ± 0.05	–	–
	1	$*0.06 \pm 0.02$	16.9 ± 15.2	0.05 ± 0.01	–	–
Sept–Oct 1975	30	0.26 ± 0.06	30.5 ± 7.2	0.21 ± 0.03	0.45 ± 0.09	0.02 ± 0.007
	31	0.23 ± 0.04	50.0 ± 9.0	0.13 ± 0.02	0.14 ± 0.03	0.01 ± 0.003
	34	0.14 ± 0.02	40.0 ± 6.2	0.13 ± 0.02	0.18 ± 0.01	–
	22	0.22 ± 0.04	58.6 ± 9.5	0.09 ± 0.02	–	–
	40	0.14 ± 0.02	41.6 ± 8.1	0.13 ± 0.02	–	–

Error, 1σ propagated.

*Ratio estimated for coastal water (ref. 17).

Table 4 Decay of separated ^{242}Cm , sample station 52 February–March 1975

Decay period (d)	^{242}Cm (fCi kg $^{-1}$)	^{241}Am (fCi kg $^{-1}$)
0	7.48 ± 0.70 (calculated)	1.48 ± 0.09
206	3.66 ± 0.47 (measured)	1.48 ± 0.09
339	1.94 ± 0.20 (measured)	1.48 ± 0.09
434	1.34 ± 0.17 (measured)	1.48 ± 0.09
628	0.63 ± 0.12 (measured)	1.48 ± 0.09
880	0.26 ± 0.04 (measured)	1.48 ± 0.09

^{242}Cm , half life = 163 d; ^{241}Am , half-life = 433 yr; $^{242\text{m}}\text{Am}$ = 152 yr.
Error, 1σ propagated.

from fallout of around 0.3×10^{-4} and 0.2×10^{-4} respectively, it is clear that these samples contain curium originating from the French fuel reprocessing plant: ratios of greater than 0.1 and 0.01 respectively were determined. Although ^{244}Cm was not detected during February–March 1975, small amounts were measured during September–October 1975. In the case of ^{242}Cm the decrease in activity was much larger than for the other actinide isotopes; the September–October activities in the Channel being about 20% of those for February–March.

The non-detection of ^{244}Cm during February–March 1975 is remarkable, especially in view of the fairly high ^{242}Cm activities in the Channel, and seems to confirm the variability, both compositionally and quantitatively, of the discharges from the French fuel reprocessing plant.

It is difficult to estimate the degree of actinide penetration eastwards into the southern North Sea. From $^{238}\text{Pu}/^{239+240}\text{Pu}$ ratios it seems that the region about $53^\circ 00' \text{N}$, $04^\circ 00' \text{E}$ is the furthest point where the source of plutonium is known to be due to waste discharge. The somewhat increased ratios of $^{241}\text{Pu}/^{239+240}\text{Pu}$ in this area compared with those reported for fallout⁶ might indicate the appearance of actinide activity from nuclear fuel reprocessing operations moving slowly eastwards.

Murray and Eicke¹⁷ have shown that although actinide activity levels along the German North Sea Coast in 1975 were mainly due to fallout, over 75% of the measured ^{137}Cs could be attributed to fuel reprocessing plants in both the UK and France. As they point out, base-line values obtained for the German North Sea and Baltic coastal waters may be valuable in showing possible changes in concentrations of actinides in these areas that might occur in the future, as well as helping to identify the source.

An attempt has been made to determine whether ^{242}Cm was supported by $^{242\text{m}}\text{Am}$. The decay activity of ^{242}Cm was measured in sample 52 taken on the February–March 1975 cruise, at intervals over a period of about 3 yr. The results are shown in Table 4. Using the procedure outlined by Dutton and Lovett¹⁸, a regression line calculated by the method of least squares gave a half life of 179 ± 25 d. This compares to a 163-d value for the

pure isotope. The fact that the 1σ error in the half life includes the standard half life of ^{242}Cm precludes the possibility of determining the existence of $^{242\text{m}}\text{Am}$. However, the establishment of its existence using the method described by Dutton and Lovett¹⁸ would require samples containing very much higher activities (to obtain small measurement errors) than those of the seawater measured in the present case. The possibility remains, nevertheless, that measurement of these samples after a period of more than 10 half lives of ^{242}Cm might reveal some residual activity.

Conclusions

The variations of activity concentrations in the English Channel between March and October 1975 have demonstrated the pulse-like nature of actinide contamination entering the southern North Sea. Comparing actinide ratios in these waters with those expected from fallout, it has been shown that the major input of the isotopes is the French nuclear fuel reprocessing plant at La Hague.

Although plutonium water activities probably only represent a small percentage (<5%) of the total released in low-level wastes, this fraction shows a surprisingly high degree of mobility. Considering the various differences in behaviour of these elements it is clearly necessary to develop methods for the analysis of their chemical forms.

Finally, as the European nuclear generating and reprocessing capacity increases over the next few decades, the level of discharge could also increase. Actinide activities in the North Sea, although at very low concentrations, may reflect this evolution in nuclear technology.

We thank Dr M. Hoppenheit for his contribution and Miss M. Domian for technical assistance.

Received 20 September 1978; accepted 6 February 1979.

1. Jefferies, D. F., Preston, A. & Steele, A. K. *Mar. Pollut. Bull.* **4**, 118–122 (1973).
2. Kautsky, H. *Di. hydrogr. Z.* **29**, 217–221 (1976).
3. Mitchell, N. *Fisheries Radiobiological Laboratory Technical Report*, **12**, 32 (1977).
4. Murray, C. N. & Kautsky, H. *Est. Coast. Mar. Sci.* **5**, 319–328 (1977).
5. Murray, C. N. & Statham, G. *Di. hydrogr. Z.* **29**, 69–75 (1976).
6. Livingston, H. D., Schneider, D. L. & Bowen, V. T. *Earth planet. Sci. Lett.* **25**, 361–367 (1975).
7. Hardy, E. P., Krey, P. W. & Volchok, H. L. *Nature* **241**, 444–445 (1973).
8. Fukai, R., Ballestra, S. & Holm, E. *Nature* **264**, 739–740 (1976).
9. Holm, E. & Persson, B. *Report Lund University L.U. FD6/(NFRA-3013)/1–7/(1977)*.
10. Livingston, H. D. & Bowen, V. T. *Environmental Toxicity of Aquatic Radionuclides: Models and Mechanisms* (eds Miller, M. W. & Stannard, J. N.) 107–130 (Ann Arbor, Ann Arbor, 1976).
11. Krey, P. W. et al. *Int. Symp. Transuranium Nuclides in the Environment*, 671–678 (IAEA, Vienna, 1976).
12. Bowen, V. T. & Livingston, H. D. *Nature* **256**, 482–484 (1975).
13. Kautsky, H. *Di. hydrogr. Z.* **26**, 241–246 (1973).
14. Hetherington, J. A., Jefferies, D. F., Mitchell, N. T., Pentreath, J. R. & Woodhead, D. S. in *Int. Symp. Transuranium Nuclides in the Environment*, 139–154 (IAEA, Vienna, 1976).
15. Livingston, H. D., Casso, S. A., Bowen, V. T. & Burke, J. C. in *XXVI Congr. and Plenary Assembly Int. Commission for Scientific Exploration of the Mediterranean Sea* (in the press).
16. Murray, C. N., Kautsky, H., Hoppenheit, M. & Domain, M. *Nature* **276**, 225–230 (1978).
17. Murray, C. N. & Eicke, H. F. *Di. hydrogr. Z.* **30**, 1–8 (1977).
18. Dutton, J. W. R. & Lovett, M. B. *Nature* **267**, 37–38 (1977).

Nerve fibre topography in the retinal projection to the tectum

J. H. Scholes

MRC Cell Biophysics Unit, University of London, King's College Biophysics Department, 26–29 Drury Lane, London WC2, UK

The orderly layout of visual fibres in the optic nerves of cichlid fishes suggests that they are guided most of the way to the brain by contact with their neighbours. Before they reach the visual map in the tectum, however, their arrangement in the cross-section of the nerve is reorganised in a way which requires longer-range guidance.

RETINAL nerve fibres in vertebrate animals grow out of the eye during development, cross the body mid-line, and spread to terminate on the surface of the opposite midbrain tectum. Their synaptic arborisations become laid out in the tectum to form an accurate neurological map of the retina, orientated in rough alignment with the visual field of the eye concerned. Surgical disturbance of the growing visual pathway has little effect on the disposition of the map which optic nerve fibre terminals

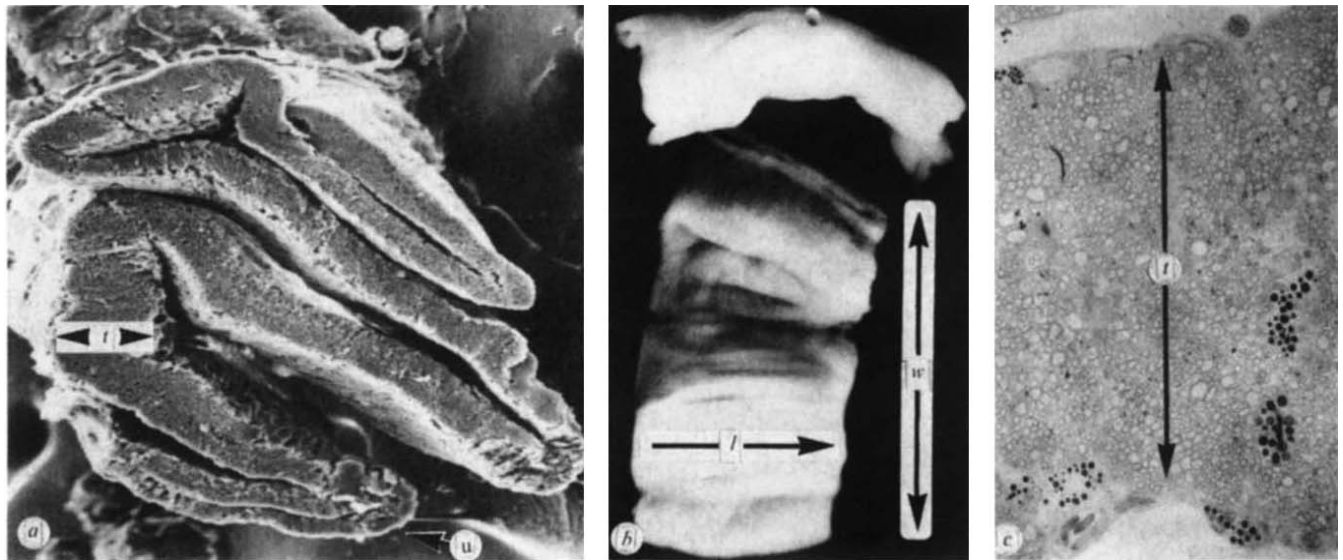


Fig. 1 *a*, Scanning electron micrograph of the cut central stump of the optic nerve in a cichlid fish, *Astronotus ocellatus*. The arrow (u) points to the unmyelinated fibre edge of the ribbon (compare Fig. 2*a*), and t is the ribbon thickness, $\sim 100\ \mu\text{m}$. *b*, Lengths of the left and right optic nerves from *Astronotus*. Both are desheathed, and the ribbon of one has been spread out flat. The arrow is the direction of fibre growth (l), and w is the width of the spread ribbon, here $\sim 7\ \text{mm}$. *c*, Transmission electron micrograph of a part of a ribbon optic nerve (*Tilapia* sp.). It comprises myelinated fibres exclusively in this region. Ribbons from different sized fish are roughly equal in thickness ($t \approx 75\ \mu\text{m}$) but vary widely in width to accommodate up to 80,000 myelinated fibres.

establish in the brain, and this finding¹ has suggested that retinal axons are guided into their synaptic array by specific affinities for predetermined loci in the tectal surface¹⁻⁴.

Nevertheless, common sense suggests that mechanical constraints to axonal growth, like neighbour contact and substrate guidance, may also contribute to the normal development of spatial order in the tectal map⁵. How they might do so is easily seen from the normal mode of retinal fibre growth within the eye. The majority of ganglion cells destined to connect with the brain accumulate serially throughout life^{6,7} in retinal tree rings, by cell division around the iris margin of the eye⁸⁻¹⁰. Thus, new optic neurites continually extend inwards from the retinal circumference to the head of the optic nerve, growing alongside their neighbours and over a template surface formed by previously differentiated axons and their satellite cells.

My purpose here is to establish whether neighbour relationships between retinal axons, laid down in the eye^{11,12}, are normally preserved throughout the remaining course of the visual projection to the tectum. To answer this question, and to understand the topological continuity of the visual pathway, I studied an informative optic nerve structure (Fig. 1*a*)^{13,14} found in certain fishes (many of the Cichlidae, and various other percoid fish). Visual axons are laid out in the cross-section of this nerve to form an accurate map of the retina¹⁴. I will show that the map is of a strikingly simple form which suggests that it is shaped by very short-range constraints (such as physical contact) to the growth of single fibres. However, at a definite point along the length of the optic nerve, almost at its central end, the map becomes topologically re-ordered in a manner which involves systematic neighbour exchanges between fibres, and therefore can only be mediated by some longer-range mechanism of guidance. I will argue that this re-ordering—the only fibre manoeuvre in the normal projection which manifestly requires long-range guidance—must also be a general feature of other lower vertebrate retinotectal pathways. Moreover, both the nature of the re-ordering, and the fact that, as discussed later, it is apparently completed before fibres arrive at the tectum, seem to preclude that its mechanism involves adhesive axonal recognition^{1,4,15} of loci within the tectal surface.

Retinotopic maps in ribbon optic nerves

Figure 1*a* shows the cut central stump of the optic nerve in a cichlid fish. Within a cylindrical outline, it contains a folded

internal ribbon of nerve fibres¹⁴, easily flattened out into a continuous sheet (Fig. 1*b*). Apart from one edge, whose relatively small cross-sectional area is filled by numerous (up to 25% of the total) closely packed fine unmyelinated fibres, the ribbon comprises myelinated axons only (Fig. 1*c*). The myelinated fibres are laid out in the cross-section of the ribbon to form an accurate map of the retinal surface, but one whose topology is not directly related to the circular outline of the eye. Instead, as shown below, the polar coordinates of the retina become laid out orthogonally in a Cartesian fibre map which is bounded by the rectangular (though folded) cross-section of the ribbon.

Figure 2*a* shows the layout, in the nerve, of (degenerating) axons from a dorsal sector of the retina, distributed as a thin continuous sheet¹⁴ throughout most of the width (w) of the ribbon, and accurately located halfway through its thickness (t). These degenerating fibres were severed inside the eye by a small cut through the retina, close to the optic disk, and parallel to the circumference of the eye, which interrupted a sector-shaped beam of peripheral dorsal retinal axons. Retinal lesions of this sort always produced sheet-like degeneration patterns extending from the unmyelinated edge of the ribbon through some fraction of its width directly related to the area of retinal surface whose fibres were severed. Thus, the w axis of the ribbon maps the retinal radius, with equal increments (Δw) representing retinal annuli of roughly equal area. The myelinated edge of the ribbon is the origin (w_0), representing the centre of the retina. This suggests that the unmyelinated fibres, numerous though they are, represent a queue waiting for consolidation into the extended myelinated ribbon of an older fish. In agreement, their degeneration patterns (detectable only by electron microscopy) extend those in the rest of the ribbon, and they comprise, as growth proceeds, a diminishing fraction of all axons in the nerve. If it is assumed that the rate of addition of new retinal cells during growth is roughly constant^{1,2}, then equal increments to the width of the myelinated part of the ribbon (for example, Δw in Fig. 2, and see below) represent equal growth periods.

Conversely, and contrary to an earlier analysis¹⁴, radial sectors of the retina are mapped in circumferential order, across the thickness (t) of the ribbon. The ribbon's two surfaces, boundaries for this linear representation of the eye circumference, correspond with a radial and ventral discontinuity in the otherwise closed outline of the retinal disk. This is the embryonic choroid fissure¹², which persists throughout adult life in

fishes with ribbon nerves. Axons from retinal cells along each edge of the fissure are guided to run along one or other surface of the ribbon (Fig. 2b).

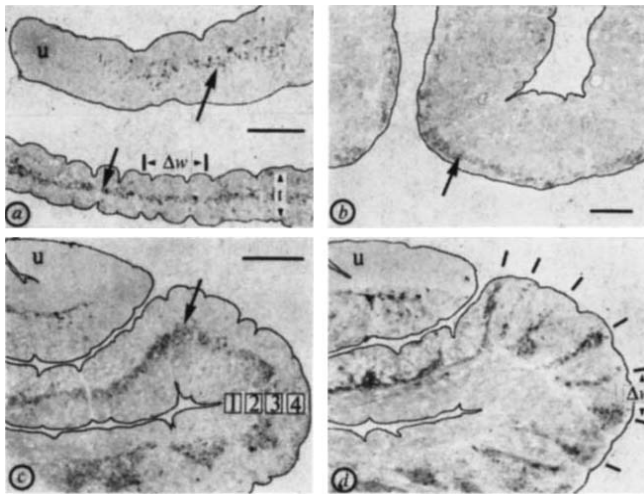


Fig. 2 Transverse sections of ribbon nerves, showing osmiophilic myelin degeneration patterns which result when populations of axons from narrow sectors of the retina are severed close to the optic disk. The sectors were mapped in animals killed after 3–4 weeks survival. Primary fixation was with aldehydes in 0.3M buffer with added divalent cations, followed by OsO_4 then epoxy embedment. The light micrographs are of unstained 2–5- μm sections, with the ribbon outlines emphasised by pen. Parts *a* and *b* are before the chiasm, and *c* and *d* beyond it. *a*, Right eye ribbon from the cichlid *Aequidens maronii* after severing a 30° sector of dorsal retinal fibres (compare positions 2–3 in Fig. 3). Degeneration debris (arrows) forms a sheet halfway through the thickness (*t*) of the ribbon, which could only be followed into the unmyelinated region (*u*) by electron microscopy at shorter survival times. Δw matches part *d* of this figure, and the scale bar is 50 μm . *b*, Similar ribbon (*Tilapia* sp.) where a retinal lesion interrupted fibres from the temporal edge of the choroid fissure (position 4, Fig. 3). Myelin debris is located along the corresponding surface of the ribbon. Scale bar, 50 μm . *c*, Part of the same ribbon as in *a*, cut further upstream almost at the left tectum, and viewed in the direction of fibre growth. Because the nerve has now wrapped around the ventral and contralateral surface of the brain, the unmyelinated edge of the ribbon (*u*) faces dorsolaterally (upwards here) rather than ventrally as at the exit from the eye (compare Fig. 3). Degenerating fibres still form a sheet. The numbered boxes define the position in the nerve of the fibre strips depicted in Fig. 4, and refer also to Fig. 3. Scale bar, 50 μm . *d*, The same ribbon approximately 70 μm closer to the tectum, at a location intermediate between the two numbered levels of the projection depicted in Fig. 4. The sheet of fibre debris has separated into blocks, one to each ribbon cross-lamina (lines), which are exchanging position with fibres at the lateral surface of the ribbon. Δw contains the same number of fibres as in *a*, showing the relationship between the fibre map grain and the size of the cross-laminae.

Changes in the outline of the fibre map

The extension of the ribbon, which I define as the ratio of width to thickness (w/t), varies along the length (l , Fig. 1b) of the optic nerve, having its smallest value ($w/t \approx 5$) where retinal axons leave the eye through the apex of the choroid fissure as a broad vertically orientated strap, not yet folded up. Extension rises to remarkable values ($w/t \approx 200$; $w > l$) in older fish as the ribbon folds up shortly behind the orbit. It condenses again to between 10 and 20 where the optic nerves cross (bodily) at the mid-line and approach the contralateral tecta. But retinal fibres remain in constant positions relative to w and t throughout, almost to the tectum, so these changes of outline are fluid deformations, only minimally affecting the local neighbour relationships of fibres within the ribbon (compare Fig. 2a and c).

In principle, the more radial transformation between circular retinal topology and the layout of axons within the nerve could arise by a process of fluid deformation similar to that involved in ribbon extension and condensation. This is most easily seen in the following way. Imagine the retinal disk is a fan, opened through 360° so that its edges meet at the position of the ventral choroid fissure. Now picture closing the fan. This manipulation compresses retinal points into an internal order similar to that assumed by corresponding fibres in the ribbon. If allowance is made for the fact that the ribbon maps the retinal radius as r^2 rather than linearly, the description below, summarised in Fig. 3, shows that axons enter the optic nerve in a projection pattern broadly isomorphous with this model transformation, which resembles a conformal mapping¹⁶.

The retinal projection into the optic nerve

In retinal whole mounts, axons of ganglion cells recently added in the periphery are seen clearly to run over the outer, vitread, surface of older established fibres, and they converge radially upon the most ventral edge of the ribbon as it leaves the eye. The more centrally the fibres originate, the more deeply they are stratified in the layer of retinal axons and the more their course is directed towards the dorsal apex of the slit-shaped optic disk. These points, true also in embryonic chick retina¹¹, are shown schematically in Fig. 3. The diagram shows how retinal fibres project into the ribbon outline without changing their local neighbour relationships; to verify that this is so, account must be taken of fibres of retinal eccentricities between the two extremes depicted (*c* and *p*).

This interpretation suggests how retinotopic order in the nerve could be generated by rather minimal physical constraints to the growth of new fibres extending from the retinal margin. Sufficient growth rules could be that individual neurites, (1) can only extend over a substrate formed by their predecessors in differentiation, and (2) cannot freely exchange neighbour relationships with others growing at the same time.

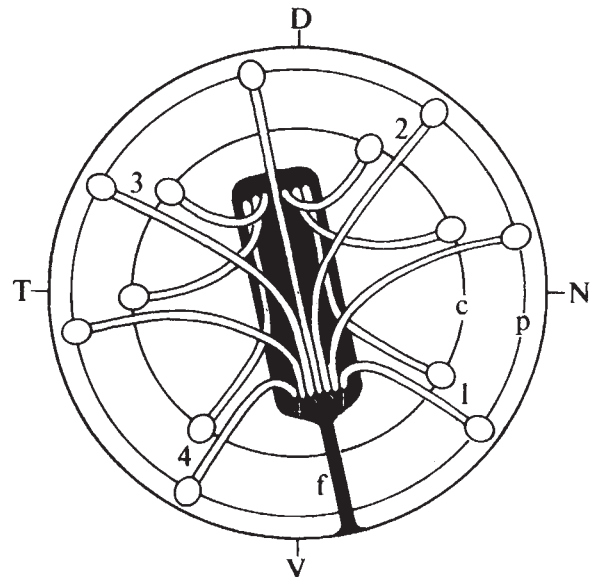


Fig. 3 A schematic picture of how fibres from an extreme peripheral annulus of retinal ganglion cells (*p*) and an extreme central one (*c*) project into the cross-section of the ribbon nerve as it leaves the right eye (viewed as if through the pupil), yet to extend and fold into the outline shown in Fig. 1a. Along each retinal radius, axons originating between these two extremes are ordered in the depth of the fibre layer according to eccentricity and take intermediate courses to the head of the nerve. Nasal, dorsal, temporal and ventral retinal quadrants are N, D, T and V respectively, and *f* is the choroid fissure. Numbered circumferential positions refer to Figs 2c and 4.

Mature patterns of axonal fasciculation in the cichlid retina and nerve, as elsewhere^{11,17}, involve frequent local exchanges between nearby fibres. If these exchanges were cumulative, it is arguable that any nascent retinotopic order should become scrambled before axons reach the brain, but degeneration patterns in the nerve show no signs of diffusing. Fasciculation patterns must, therefore, represent entirely local segregations of fibres into bundles—a kind of random bracketing which could well be mediated by glial ensheathment and continued tissue growth.

How fibres leave the ribbon

Cartesian retinotopic order persists throughout almost the entire length of the ribbon (Fig. 2a, c), to the point where its outlines begin to merge into the marginal fibre brachia of the contralateral tectum. This prompted the question whether considerations shaping the form of the fibre map also play a part in the accurate assembly of optic axon terminals into the visual map on the tectum itself. First of all, does the fibre map in the ribbon unfold again conformally like a fan, during radiation onto the tectum, as it became folded (Fig. 3) at the optic nerve head? The short answer is no.

Just before (150–300 μm) the folded outline of the ribbon coalesces into the tectal brachia, its entire width (w) sub-divides into a series of cross-laminae, each comprising a few hundred fibres. The laminae have very sharp boundaries, visible in thick sections of normal nerves as shearing planes which separate countercurrents of fibres running extremely obliquely in opposite directions across the thickness (t) of the ribbon. Within laminae, however, fibres change their positions in the ribbon by more fluid movements, for the most part preserving neighbour relationships.

Particularly for axons at the lamina boundaries, these manoeuvres abruptly disrupt neighbour relationships between fibres, otherwise preserved throughout the entire retinotectal projection. Their major effect is to re-order, piecewise, the fibre map of the eye circumference laid out across the t axis of the ribbon. The movements are illustrated in Fig. 2c and d, showing how degenerating fibres from a dorsal retinal lesion, first ordered in a continuous sheet (Fig. 2a, c), separate into blocks and move onto the lateral surface of the tract (Fig. 2d). Figure 4, compiled from similar data for other retinal sectors, shows schematically how all fibres within a single cross-lamina (representing a retinal annulus) become re-ordered by these exchanges, which were identified many years ago in the roach¹⁸.

Because axons in each lamina of the ribbon move independently in the same way, the outcome is to reverse the handedness of the entire fibre map, and to re-allocate its boundaries to a new point on the eye circumference. Handedness here means the direction, relative to the growth axis (w), in which the eye circumference is mapped across the thickness of the ribbon (t). To specify handedness, a viewpoint must be defined, and the natural one, adopted in Figs 2 and 4, is the direction along which fibres grow and conduct impulses. As visual fibres originate on the outer surface of the retina, this convention dictates that the layout of ganglion cells be viewed, as in the keys to Fig. 4, from behind the eye. Thus, Fig. 4 shows that, in an otherwise continuous projection, a coarse-grained reversal of map handedness at the cross-laminae enables fibres to proceed to their tectal loci without further exchange of their neighbour relationships.

Irrespective of retinal eccentricity, established myelinated fibres complete their re-ordering together within a very short distance (for example, 200 μm) compared with the pathway length (for example, 1 cm) accrued by bodily growth since the first embryonic fibres arrived at the tectum. Thus, although it was not possible to trace, by degeneration methods, the small minority of unmyelinated fibres engaged in growth at any instant, extrapolation suggests they also re-order at the same level.

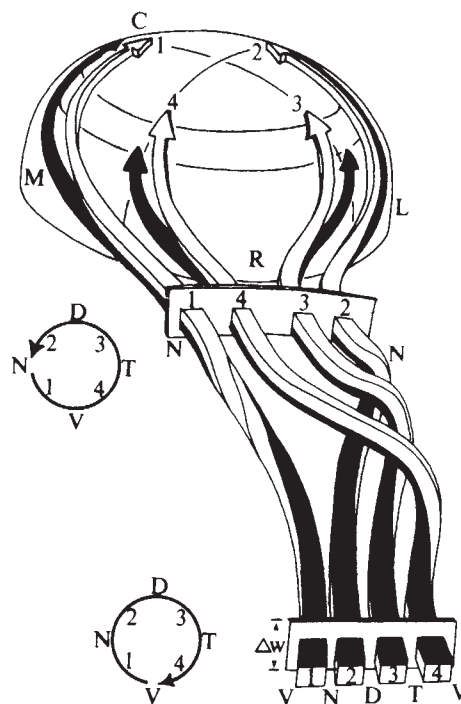


Fig. 4 Schematic interpretation of how retinal axons within a single cross-lamina of the ribbon nerve become internally re-ordered as they approach the rostral (R) pole of the left tectum. The strips represent blocks of axons from different sectors of an annulus of ganglion cells in the right retina, numbered and labelled as in Fig. 3. Their starting configuration corresponds with the numbered blocks superimposed on the micrograph of Fig. 2c. Fibre populations in each cross-lamina of the ribbon (compare Fig. 2d) re-order independently and in the same way. Each maps an increment (Δw , Fig. 2a and d) of the retinal radius. Accordingly, the black surfaces of the strips are fibres most recently added to the cross-lamina, depicted rather conjecturally as they could not be traced fully. The course of the strips onto the tectal map is consistent with the fibre radiation pattern seen in whole mounts, although the twist handedness depicted for strip 1 is arbitrarily chosen. The keys on the left are back-projections of the lamina to its origin in the right eye (shown as if viewed from behind) to show how re-ordering reverses the circumferential sequence (arrows) of the fibre map, and re-allocates its boundaries (gaps) to a nasal position. M, C, L, and R are medial, caudal, lateral and rostral poles of the tectum respectively; other labels as in Figs 2 and 3.

Other retinotectal projections

It is important to realise that the cross-laminar fibre re-ordering in the cichlid ribbon is not related to any basic mismatch of handedness between retinal image and tectal map. They are, as in other lower vertebrates, superimposable when properly viewed from a single frame of reference, for example the standpoint of axonal growth. It relates instead to a general mismatch between the orientation of the tectal map and the direction of optic tract approach. This is best understood from Fig. 4 by noting that the positional exchanges depicted would be unnecessary were the pathway to approach the tectum medially rather than rostrally.

Because the mismatch is a consequence of the relative layout of the tract and the tectum, it must be a general characteristic of other lower vertebrate retinotectal pathways, which the cichlid projection resembles in all particulars except the unusual ribbon structure of its optic nerve. Data in the literature make it possible to state this argument more explicitly. In amphibia, the axis along which new fibres accrue to the optic tract²¹, and the direction through which the tract approaches the tectum, are the same as in the cichlid pathway at a comparable level. Moreover, orthogonal to the growth axis of the optic tract, amphibian visual fibres are laid out in a retinal circumferential sequence (NVTDN, Fig. 4)²² matching that into which cichlid axons

emerge from the ribbon cross-laminae (Fig. 4), similarly bounded and of the same handedness. As the arrangement is of opposite handedness to the layout of retinal fibres as they originate in the eye (Fig. 4, keys), some reorganisation of the fibre map similar to that achieved in the cichlid cross-laminae must already have been mediated before the amphibian optic tract emerges onto the lateral wall of the diencephalon, possibly as far downstream as at the optic chiasm. Indeed fibre interdigitation patterns in fused chiasms have a certain resemblance of appearance to the cross-laminae located further upstream in the visual pathway of cichlids, whose optic nerves, as in fish generally, cross bodily at the mid-line.

Conclusions

It is hazardous to infer prior developmental mechanisms from differentiated structures. Within their limits, however, these observations allow certain useful conclusions. First, an extremely short-range mechanism, possibly no more sophisticated than physical contact between adjacent retinal axons, is evidently consequential enough to determine the layout of optic nerve fibres throughout most of their course to the tectum, albeit a layout incompatible with that which their axon terminals must adopt in the tectal map. Second, the topological re-ordering necessary for axons to reach their correct terminal sites in the tectum involves neighbour exchanges and must, therefore, be

guided by longer-range constraints to fibre growth. This re-ordering, which I have argued must be a general feature of lower vertebrate retinotectal projections, seems to be mediated, particularly in Amphibia, some distance before growing fibres reach the tectum. It is unlikely, therefore, to be the outcome of exploratory axonal growth towards putative neuronal recognition sites within this tissue. Finally, the piecewise nature of the re-ordering means that the number of retinal fibres which exchange their neighbours is, in fact, a small fraction of the total. It can be seen, therefore, that the mechanism which guides re-ordering of the fibre map needs in principle only to act on a fraction of the fibres which are involved in the movements; the rest might follow by neighbour guidance.

Current understanding of the ontogeny of the retinotectal projection is that the accretion of new retinal fibres is a continuous process^{7,23,24}. One urgent question raised by the findings described here is how the continual growth of the pathway is punctuated to enable the large blocks of fibres, comprising single cross-laminae (Δw , Figs 2 and 4), to reorganise as independent units.

I thank the Electron Microscopy Unit at Queen Mary College, London, and particularly Dr P. Mobbs, for making scanning micrographs and Dr W. Lieb, Professor B. B. Boycott FRS, Dr D. S. Gilbert, and Christine Holt for critically reading the manuscript, and for their helpful suggestions.

Received 27 November 1978; accepted 22 February 1979.

1. Sperry, R. W. *Proc. natn. Acad. Sci. U.S.A.* **50**, 703–710 (1963).
2. Gaze, R. M., Jacobson, M. & Szekely, G. *J. Physiol., Lond.* **165**, 484–499 (1963).
3. Jacobson, M. *Dev. Biol.* **17**, 202–218 (1968).
4. Hunt, R. K. & Jacobson, M. *Curr. Topics dev. Biol.* **9**, 203–259 (1974).
5. Cook, J. E. & Horder, T. J. *Phil. Trans. R. Soc. B278*, 261–276 (1977).
6. Müller, H. *Zool. Jb.* **63**, 275–324 (1952).
7. Johns, P. R. & Easter S. S. *J. comp. Neurol.* **176**, 331–342 (1977).
8. Hollyfield, J. G. *J. comp. Neurol.* **144**, 373–380 (1972).
9. Strasnick, K. & Gaze, R. M. *J. Embryol. exp. Morph.* **26**, 67–69 (1971).
10. Johns, P. R. *J. comp. Neurol.* **176**, 343–358 (1977).
11. Goldberg, S. & Coulombre, A. J. *comp. Neurol.* **146**, 507–518 (1972).
12. Bunt, S. M. & Horder, T. J. *J. Physiol., Lond.* **272**, 11P–12P (1977).
13. Polyak, S. *The Vertebrate Visual System* (University of Chicago Press, 1957).
14. Anders, Juanita J. & Hibbard, E. *J. comp. Neurol.* **158**, 145–154 (1974).
15. Marchese, R. B., Barbera, A. J. & Roth, A. *Ciba Fdn Symp.* **29**, 315–327 (1975).
16. Churchill, R. V. *Complex Variables and Applications* (McGraw-Hill, London, 1960).
17. Maturana, H. R. *J. biophys. biochem. Cytol.* **7**, 107–120 (1960).
18. Lubsen, J. *Ned. T. Geneesk.* **67**, 1258–1260 (1921).
19. Horder, T. J. *Brain Res.* **72**, 41–52 (1974).
20. Attardi, D. G. & Sperry, R. W. *Expt Neurol.* **7**, 46–64 (1963).
21. Gaze, R. M. & Grant, P. J. *Embryol. exp. Morph.* **44**, 201–216 (1978).
22. Scalia, F. & Fite, K. J. *comp. Neurol.* **158**, 445–478 (1974).
23. Gaze, R. M., Keating, M. J. & Chung, S. H. *Proc. R. Soc. B185*, 301–330 (1974).
24. Jacobson, M. *Brain Res.* **127**, 55–67 (1977).

letters

The accuracy of predicted radioastronomical frequencies and the spectrum of hydroxyl

IF radioastronomers are to use predicted frequencies to search for new interstellar species, they require some fairly precise indication of the accuracy of the theoretical calculations. Calculations of rotational constants are not very accurate by the standards of radiofrequency spectroscopy, but computations of Λ -doublet splittings have seemed remarkably accurate^{1–3}. However, these calculations were presented without error limits and as it is not possible to produce these from pure theory we have used the hydroxyl radical to probe the limits of accuracy. The experimental frequencies are known from observations^{4,5} of the interstellar maser action in OH to a high precision, and it is a particularly suitable case for computation. We show here that the prediction for the lowest rotational level is accurate to 1 MHz; an accuracy higher than conventional terrestrial experimental spectroscopy.

By considering Λ -doubling of rotational levels of $^2\Pi$ states to be caused by second order perturbations, Van Vleck⁶ showed

that the splitting may be expressed in terms of the parameters $(\frac{1}{2}p + q)$ and q :

$$\Delta\nu = -(\frac{1}{2}p + q)[1 \pm (2 - Y)/X](J + \frac{1}{2}) \mp (2q/X)(J + 3/2)(J + \frac{1}{2})(J - \frac{1}{2})$$

Here the upper signs refer to F_2 sublevels and the lower to F_1 ;

$$X^2 = Y(Y - 4) + 4(J + \frac{1}{2})^2 \quad \text{and} \quad Y = A_{\pi}/B_{\pi}$$

The parameters p and q depend on integrals over vibronic wave functions of the operators $\hat{B}(\hat{L}^+ + \hat{L}^-)$ and $\hat{H}_{SO}$. These matrix elements are factorised into vibrational and electronic integrals so that the contribution of a single $^2\Sigma^+$ electronic state to the splitting in the ν th vibrational level of a $^2\Pi$ state is given by

$$p_{\nu} = 4\langle^2\Pi|\hat{H}_{SO}|^2\Sigma\rangle\langle^2\Pi|\hat{B}(\hat{L}^+ + \hat{L}^-)|^2\Sigma\rangle \times \sum_{\nu'} \frac{\langle\nu|\nu'\rangle\langle\nu|\hat{B}|\nu'\rangle}{E_{\pi,\nu} - E_{\Sigma,\nu'}}$$

and

$$q_{\nu} = 2\langle^2\Pi|\hat{B}(\hat{L}^+ + \hat{L}^-)|^2\Sigma\rangle^2 \sum_{\nu'} \frac{|\langle\nu|\hat{B}|\nu'\rangle|^2}{E_{\pi,\nu} - E_{\Sigma,\nu'}}$$

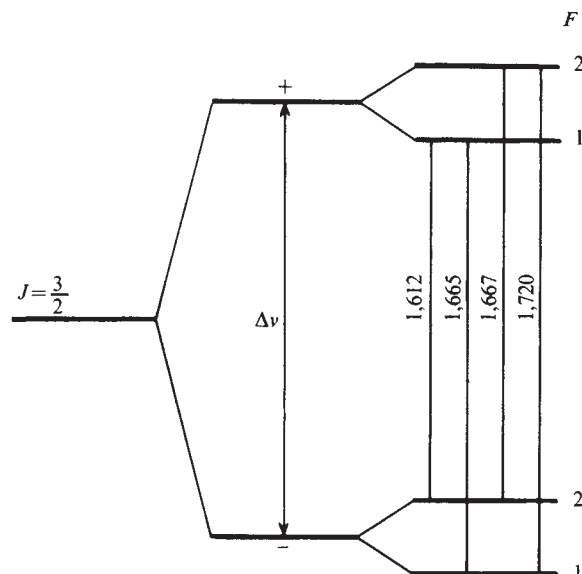


Fig. 1 Λ -doubling ($\Delta\nu$) in the lowest rotational level of OH ($^2\Pi_{3/2}$; $J=3/2$) showing the observed radioastronomical transitions in MHz (not to scale).

where the summation should be taken over all vibrational levels, v' . Contributions from all $^2\Sigma$ states are then added with $^2\Sigma^-$ states giving negative terms.

An earlier study¹ calculated p and q for OH assuming that molecular orbitals were invariant between electronic states and restricted summations by assuming Van Vleck's case of pure precession. Here we have calculated separate wave functions for $X^2\Pi$ and $A^2\Sigma^+$ using Cade and Huo's atomic orbital basis set⁷ in the ALCHEMY program⁸. The electronic functions were then improved by configuration interaction (CI) using 15 Π configurations and 11 of Σ symmetry. This was done in stages to ensure that a limit had been achieved; further additions did not change the computed values of p and q .

Morse functions were used for vibrational wave functions after showing that these give identical results to those obtained with numerical solutions from RKR curves. The vibrational summations over v' extend to $v'=5$; terms for $v'=5$ and above do not contribute significantly.

The effects of $B^2\Sigma^+$ were calculated to be less than those of $A^2\Sigma^+$ by a factor of about 10^9 , indicating how good Van Vleck's pure precession approximation is in this case.

Figure 1 indicates the observed astronomical transitions with epicentres separated ($\Delta\nu$) by 1,666 MHz. Table 1 shows calculations of p and q and of this separation.

This lowest rotational level is the one of prime importance astrophysically and the agreement between theory and experiment is essentially exact, being better than 1 MHz. However, the splitting in this particular level depends mostly on the value of the constant q . A slightly more difficult case for theory is provided by the manifold excited rotational states, where p as well as q has an influence and errors are magnified by multiplication of the constants by higher powers of the rotational quantum number.

Table 1 Λ -doubling in the $X^2\Pi$ ($v=0$; $J=3/2$) state of OH

	p (cm^{-1})	q (cm^{-1})	Splitting (MHz)
Hinkley <i>et al.</i>	0.242	-0.0391	
This work without CI	0.2308	-0.03807	
This work with CI	0.2378	-0.03982	1,666
Terrestrial expt	0.2348	-0.0387	1,654
Astronomical expt			1,666

For $J=\frac{1}{2}$ of $^2\Pi_{1/2}$ the separation of epicentres of the observed astronomical lines is 4,730 MHz, for which our calculations predict, 4,795 MHz; $J=5/2$ of $^2\Pi_{3/2}$ has a terrestrially measured separation of 6,033 MHz and we calculate 5,946 MHz, whereas the even more highly excited level $J=5/2$ of $^2\Pi_{1/2}$ has an epicentre separation of 8,153 MHz and our result is 8,444 MHz.

These results suggest that the realistic limits for our calculations, if done with care, are a few megahertz. This is better than conventional spectroscopy. We intend to recompute the Λ -splitting in SiH to this accuracy so that calculations of this type can be used to predict as yet unobserved interstellar molecules and ions.

The computational work was supported by a grant from the SRC and forms part of a project supported by a NATO research grant.

D. L. COOPER
W. G. RICHARDS

Physical Chemistry Laboratory,
South Parks Road,
Oxford, UK

Received 18 January; accepted 6 March 1979.

1. Hinkley, R. K., Walker, T. E. H. & Richards, W. G. *Proc. R. Soc. A* **331**, 533-560 (1970).
2. Hammersley, R. E. & Richards, W. G. *Nature* **251**, 597-598 (1974).
3. Wilson, I. D. L. & Richards, W. G. *Nature* **271**, 137 (1978).
4. Weinreb, S., Barrett, A. M., Meeks, M. L. & Henry, J. C. *Nature* **200**, 829-831 (1963).
5. Turner, B. E. & Heiles, C. E. *Astrophys. J.* **170**, 453-462 (1971).
6. Van Vleck, J. H. *Phys. Rev.* **33**, 467-506 (1929).
7. Cade, P. E. & Huo, W. M. *J. chem. Phys.* **47**, 614-618 (1967).
8. Bagus, P. E. in *Proc. Seminar Selected Topics in Molecular Physics* (IBM Germany, Ludwigsburg, 1969).

A nonspherical thermal instability in young stars

NONSPHERICAL thermal instabilities have been shown¹⁻⁴ not to be excited during the hydrogen-burning phase of stellar evolution. However, this conclusion holds true only for models with a homogeneous distribution of heavy elements. We show here that in nonhomogeneous young stars more massive than, or of the same mass as, the Sun, large-scale nonspherical thermal instability can be excited during the short period of time when carbon just starts capturing protons but before convection has set in.

The non-homogeneous structure is expected to form in the case of a dusty protocloud. In this case gravitation can collect the dust at the centre of gravity of the cloud^{5,6} or the dust can fall into the centre during cloud contraction⁷. We assume here that, although heavy elements are concentrated near the centre, their mass fraction (Z) is small everywhere and Z is proportional to Z_{12} , the mass fraction of carbon. Convective mixing of the central region before the ignition of nuclear reactions is assumed not to occur. This is true if the mass of the star is sufficiently large⁸⁻¹⁰. For simplicity, effects due to rotation and radiative pressure are ignored in the following.

For thermal instability to occur ϵ , the rate of nuclear energy generation, must be highly sensitive to T , the temperature. We take ϵ to be proportional to $\rho X Z_{12} T^\nu$, where ν is assumed to be constant. The dependence of ϵ on Z_{12} makes this sensitivity greatly enhanced in our case of a small Z , that is, of large h_μ/h_Z , where μ is the molecular weight, while scale height $h_\mu = -(\partial \ln \mu / \partial r)^{-1}$, and so on. (Here $h_\mu > 0$, and $h_Z > 0$, but h_X is likely to be negative.) In other words, for this rapidly growing instability any displacement leads to a change in mass fraction of heavy elements which is much larger than, and of the same sign as, the change in molecular weight. Furthermore, the excitation of the nonspherical thermal instability leads to negligibly small distortions in pressure and density distributions, we find¹⁻⁴ that

$\mu^*/\mu = T^*/T$, that is, the relative eulerian perturbations of molecular weight and temperature are equal. Thus, provided that the time of instability growth is much shorter than the timescale of change of the unperturbed energy generation rate, we find that a large increase (decrease) in this rate corresponds to a small increase (decrease) in the temperature. This conclusion follows from an equation^{2,4} generalised to include variations in $Z_{12} = \text{const}$. Z :

$$\varepsilon^* T(\varepsilon T^*)^{-1} = \nu_{\text{eff}} = \nu + h_\mu(h_Z^{-1} + h_X^{-1})$$

As $h_X = h_\mu$ and $h_\mu \gg h_Z$, we have $\nu_{\text{eff}} \approx h_\mu/h_Z$. Usually, $\nu \approx 15-20$, while Richstone's⁴ estimate shows that some mean value of $\nu_{\text{eff}} \approx 20$ makes the nonspherical thermal instability likely. If Z equals a few per cent then $\nu_{\text{eff}} \approx 100$. Such a value of ν_{eff} is so large that the probability of instability excitation is high even though, in our contracting model, most of the energy liberated is due to the contraction of the model. This instability excitation certainly occurs in models with a sufficiently large mean value of ν_{eff} , in which case the perturbed opacity can also produce a destabilising effect. For the most rapidly growing perturbation, the point of the highest temperature is displaced while the distributions of pressure and density remain almost unaltered.

Here we are considering the phase of evolution before the formation of the convective core; this means that f , the radial component of unperturbed entropy gradient, decreases and tends to zero. Therefore, rapid motions should be excited to ensure the necessary change in entropy to fulfil the energy balance in the perturbed state, which is characterised by almost the same entropy distribution as the unperturbed state. In other words, the time of growth can be very short in the model under consideration^{2,4}. The minimum value of the time of growth (τ_{min}) is determined by when perturbations of pressure and density fields become significant. This gives $\tau_{\text{min}} \approx (r^2 \varepsilon^{-1} h_p h_Z h_*^{-2})^{1/3}$, where $h_* = |\nabla \ln p^*|^{-1}$. It is easy to take the time dependence of τ into account if f can be modelled by a linear function of time.

Detailed computation is necessary to find that mean value of h_μ/h_Z above which a rapidly growing instability is really excited in the model chosen. When $\nu_{\text{eff}} \approx 100$, the instability apparently does exist in many models. Using Iben's¹¹ models of contracting stars we find that time τ_{min} is usually of the order of 0.1 yr. As the molecular weight of each moving elementary volume is conserved, we expect, as a result of the instability, that large-scale relative spatial variations $T^*/T \sim \mu^*/\mu \sim h_p/h_\mu$ are formed, whereas $v_r^* \sim h_*/\tau$, and $p^* \leq r^2 \rho/\tau^2$; for a $3M_\odot$ model, $v_r^* \sim 10^4 \text{ cm s}^{-1}$, and $p^*/p \leq 10^{-7}$, the nonspherical motions produced by the instability are large enough to transfer all the heat produced by the energy source. Also, these motions can effectively amplify the magnetic fields which are expected to be present in young stars. A field-induced instability followed by emergence of frozen-in fields on the surface can be responsible for the erratic activity usually observed in young stars.

Considering outbursts in FU Ori type stars¹², an interesting model would be one undergoing recurrent phenomena involving a temporal increase of the energy generation rate by more than two orders of magnitude on a scale of some months. If this increase were caused by the nonspherical thermal instability excited at the moment when the equilibrium is nearly adiabatic, then the time of growth can be as short as a month. However, to ensure that ε increases we have to accept that the instability can lead to the formation of a nonspherical structure containing points at which the temperature is up to 1.5 times higher than the unperturbed temperature. In this case the unperturbed model must have a large central concentration of heavy elements. More thorough study is needed to clarify whether such a model can be unstable. If the instability is possible, and its excitation produces large variations in the temperature field, the pressure field will also be disturbed. In such circumstances local convection is likely in some regions with the convective heat flux depending on the horizontal coordinate. If periods of liberation of considerable energy are followed by the energy source zone

expanding slightly at the same time as local convection reduces due to the horizontal temperature differences, then the contraction phase which was terminated by the instability excitation can resume. In this rather speculative sequence of recurrent outbursts, the local convection (perhaps coupled with the magnetic field induced instability) is supposed to be effective in transporting heat but not in causing full mixing of the energy source zone.

Nuclear reactions seem to maintain and amplify the horizontal temperature variations of the type in question¹³. That is because the ratio T/μ remains spherically symmetric, the temperature and hence the nuclear burning rate are higher within those regions where the burning element is more exhausted and, hence, the molecular weight is larger. Thus, we cannot exclude the possibility that the structure with the horizontal temperature variations can survive the stage of hydrogen burning at least in models not having convective cores, or in models where the convection is modified (in the presence, for example, of a magnetic field) to such an extent that the convective mixing does not destroy the horizontal variations mentioned. I have suggested elsewhere that a similar state is realised in magnetic stars¹³, in which case, the dimensions of the model seem to be larger and, hence, the rotation to be slower. These facts apparently agree with the observational data¹⁴.

I thank the N. Copernicus Astronomical Center for hospitality, also E. Ergma and B. Paczyński for helpful discussions and D. S. Hall for language corrections.

YU. V. VANDAKUROV

N. Copernicus Astronomical Center,
Polish Academy of Sciences,
00-716 Warsaw, Poland

Received 8 January; accepted 23 February 1979.

Permanent Address: A. F. Ioffe, Physical Technical Institute, Academy of Sciences of USSR, Leningrad, 194021, USSR.

1. Mestel, L. *Astrophys. J.* **126**, 550-558 (1957).
2. Kippenhahn, R. Z. *Astrophys. J.* **67**, 271-282 (1967).
3. Rosenbluth, M. N. & Bahcall, J. N. *Astrophys. J.* **184**, 9-16 (1973).
4. Richstone, D. O. *Astrophys. J.* **188**, 327-333 (1974).
5. Horedt, Gp. *Astrophys. Space Sci.* **45**, 353-367 (1976).
6. Alfvén, H. & Carlqvist, P. *Astrophys. Space Sci.* **55**, 487-509 (1978).
7. Harichev, I. A. *Sci. Inform. (USSR)* **33**, 79-88 (1974).
8. Larson, R. B. *Mon. Not. R. astr. Soc.* **145**, 271-295 (1969).
9. Larson, R. B. *Mon. Not. R. astr. Soc.* **157**, 121-145 (1972).
10. Westbrook, C. K. & Tarter, C. B. *Astrophys. J.* **200**, 48-60 (1975).
11. Iben, I. Jr. *Astrophys. J.* **141**, 993-1018 (1965).
12. Herbig, G. H. *Astrophys. J.* **217**, 693-715 (1977).
13. Vandakurov, Yu. V. *Letters A.J. (USSR)* **4**, 362-366 (1978).
14. Preston, G. W. in *Stellar rotation* (ed. Slettebak, A.), 254-262 (Reidel, Dordrecht, 1970).

Drifting spikes—a new species of solar radio bursts

OBSERVATIONS of solar radio bursts of spectral type IV have revealed the presence of a complicated fine structure in the emission at metric wavelengths¹⁻⁴. It would be interesting to extend high resolution spectral observations to decimetric wavelengths to explore the structure of radiation which is generated at lower heights in the solar atmosphere, and is, therefore, closer to the site of the associated optical flare. We report here a new species of fine structure bursts, drifting spikes, which have been observed in the 500 MHz frequency region.

A radio spectrograph covering frequencies $490 \leq f \leq 550 \text{ MHz}$ has been in operation at the Oslo Solar Observatory since 1972. The instrument sweeps across the observed band at a rate of 50 times per s, and has a frequency resolving power of 0.3 MHz.

Drifting spikes were recorded on 18 September 1974, 11.14-11.17 UT, 23 September 1978, 10.14-10.18 UT and 27

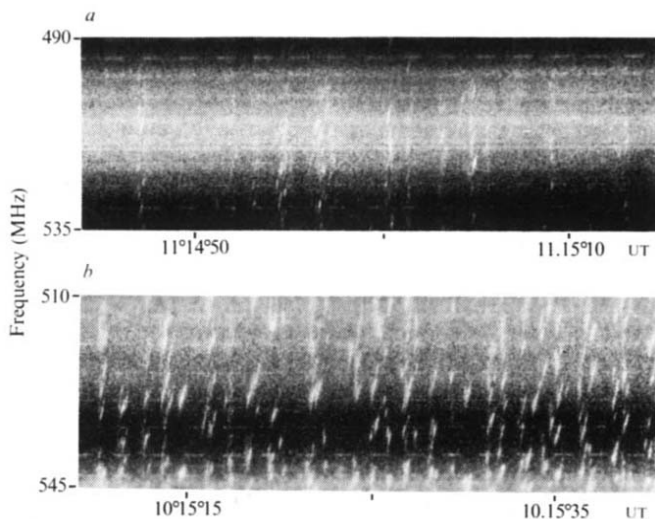


Fig. 1 Dynamic spectra showing drifting spikes observed at: *a*, 18 September 1974; *b*, 23 September 1978. Time marks, each lasting for 1 s, are seen as weak horizontal dashes. They occur at frequencies separated by 5 MHz.

September 1978, 14.45–14.47 UT. In all cases type IV bursts were in progress and flares of importances 2N, 2B and 2B showing two-ribbon structure were observed in H α at the respective times.

Simultaneous observations with radio spectrographs in the bands 140–170 MHz and 310–340 MHz revealed no drifting spikes. But bursts of this type were present on records obtained with the Dwingeloo radio spectrograph on 23 September 1978. The instrument was then tuned to cover the bands 509–533 MHz, 574–614 MHz and 655–666 MHz. The phenomenon may thus be typical at decimetric wavelengths. It is not known whether drifting spikes appear at even shorter wavelengths.

Two examples of drifting spikes are given in Fig. 1 which shows that a large number of tiny bursts are superimposed on a smooth continuum emission. Each elementary burst has a short duration, narrow bandwidth and a frequency drift from higher to lower frequencies.

Several parameters typical of individual bursts were measured on the record from 23 September 1978, which contained the

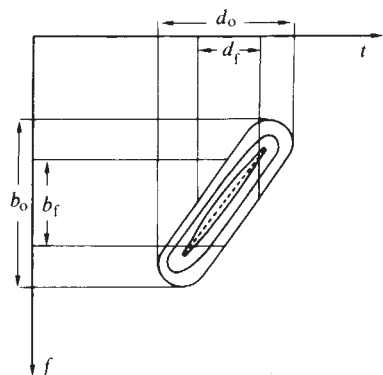


Fig. 2 The definition of some relevant burst parameters in the time-frequency plane: b_f , instantaneous bandwidth; b_o , frequency range (total range of frequency occupied by the burst at all times); d_f , duration at one frequency; d_o , lifetime (total duration of the burst at all frequencies). The slope of the burst is indicated by a dotted line, from which the frequency drift rate, df/dt , is determined.

most spectacular event. (Denotations in the following are as in Fig. 2.) The frequency range covered by single bursts (b_o) was small, the average value being 2.7 MHz. It was difficult to determine the instantaneous bandwidth b_f with good precision. Only an upper limit, $b_f \leq 1.5$ MHz can be determined. Thus drifting spikes are very monochromatic, $b_o/f = 0.5\%$ and $b_f/f \leq 0.3\%$.

For the lifetime, d_o , the average value was 0.12 s. Another quantity of interest is the duration at one frequency (d_f), but as most bursts were not adequately resolved with our sweep repetition rate of 50 Hz, it is concluded that $d_f \leq 50$ ms.

All spikes showed a frequency drift. The average value of df/dt was -23 MHz s^{-1} , giving $(1/f)(df/dt) = -0.04 \text{ s}^{-1}$. This is larger than, but of the same order of magnitude as, that found for intermediate drift bursts (fibre bursts) in the same frequency range. The scatter in burst slopes was surprisingly small, and positive drifts were not seen.

Drifting spikes frequently appeared as pairs at roughly the same frequency with the elements separated in time by 0.15 s. There was also a second and very distinctive type of ordering of the bursts. On the records they formed columns because drifting spikes were generated almost simultaneously at many frequencies separated by intervals of 5–10 MHz. This was demonstrated even more clearly by a comparison between our records and those obtained with the radio spectrograph at Dwingeloo. The effect is similar to that seen when stria bursts are ordered into decametric-type IIIb bursts.

Another important property was a regular repetition of the columns of drifting spikes. This is seen in Fig. 1b. The drifting spikes are generated at time intervals which have an average value of 1.2 s. It is interesting that before the occurrence of drifting spikes, the continuum intensity showed oscillations (pulsating structure).

If the emission frequency is equal to, or near to the plasma frequency, the occurrence of drifting spikes between 500 and 666 MHz implies electron densities $N_e = 3\text{--}5.5 \times 10^9 \text{ cm}^{-3}$ at the place of generation. New observations at higher frequencies may lead to an increase of the upper limit. The quoted electron densities correspond to heights below 50,000 km in active regions. Single frequency durations $d_f \leq 50$ ms are of the same order of magnitude as the electron collision time in these regions.

The phenomenon described here is of particular interest in connection with other observations of decimetric and centrimetric fine structure^{5–7} and a theoretical discussion of mechanisms for the production of millisecond bursts. A possible quasi-periodic generation of such bursts by plasma turbulence in flare regions has been envisaged⁸.

Further explorations of the finest structures in solar radio bursts at short wavelengths may lead to the detection of basic, small scale acceleration processes in solar flare regions.

A comprehensive analysis of drifting spikes will be given elsewhere.

We thank Dr C. Slottje for the opportunity to inspect the Dwingeloo radio spectrograph records, also T. Engebretsen and B. Rygh at Oslo Solar Observatory.

Ø. ELGARØY
O. P. SVEEN

*Institute of Theoretical Astrophysics,
Oslo University,
Blindern, Norway*

Received 14 December 1978; accepted 1 March 1979.

1. Elgarøy, Ø. *Nature* **184**, 887–888 (1959).
2. Elgarøy, Ø. *Astrophys. Norv.* **7**, 123–263 (1961).
3. Slottje, C. *Solar Phys.* **25**, 210–231 (1972).
4. Chernov, G. P. *Sov. Astr.* **20**, 449–459 (1976).
5. Elgarøy, Ø. & Sveen, O. P. *Solar Phys.* **32**, 231–240 (1973).
6. Drøge, F. *Astr. Astrophys.* **57**, 285–290 (1977).
7. Slottje, C. *Nature* **275**, 520–521 (1978).
8. Kuijpers, J. *Astr. Astrophys.* **69**, L9–L12 (1978).

Detection of circularly polarised light from noctilucent clouds

NOCTILUCENT clouds are very tenuous clouds that appear in the high latitude mesosphere during the summer¹. They are seen by scattering of sunlight from cloud particles near the mesopause. They are known to be at altitudes of ~82 km, more than 50 km higher than the next highest clouds in the atmosphere, the nacreous or mother-of-pearl clouds. They are usually seen after midnight in June and July in the north-east twilight arch at latitudes of 55–65° and appear silvery-blue² in the red twilight glow. Observations have been made from ground level^{3,4}, rocket⁵ and satellites^{6,7}. Studies of sunlight scattered from the clouds have concentrated on measurements of polarisation of the scattered light to allow estimates of the size distribution of the cloud particles to be made. These estimates agree in that only submicron particles with an upper radius limit in the range 0.05–0.3 μm are present. Recent measurements⁴ of scattered light have shown the presence of a small proportion of circular polarisation, indicating that the cloud particles are probably non-spherical, and estimates of the particle size, being based on the assumption that scattering is from homogeneous spheres (Mie scattering), may need revision. The revision might be enough to force reappraisal of the amount of water condensed to form the clouds and consequent reappraisal of the relative humidity in this part of the atmosphere. A small amount of circular polarisation can be produced in the scattering of natural (unpolarised) light from elongated particles⁸ but a greater proportion of circular polarisation will arise if the incident light is partly polarised and if the axes of the elongated particles show some directional preference. However, light scattered from spherical particles will also show some small amount of circular polarisation, if the incident light were to be linearly polarised in a direction skewed with respect to the plane of scattering. If the circular polarisation is produced by elongated cloud particles, it is possible that polarisation of the scattered light will show marked differences from place to place due to different proportions and directions of alignment of particles being caused by

updraughts or vertical wind shears present in different parts of the cloud region. This would not be expected if the circular polarisation was due to scattering of linearly polarised light from spherical particles. Apart from early photographic observations³, recent measurements of the polarisation in noctilucent clouds have been photo-electric and there has been no simultaneous measurement at different points spread over an appreciable part of the cloud display. We have developed and report here a method which allows us to test how circular polarisation varies over the area of display⁹.

A fixed-polariser/rotating-retardation plate assembly was placed in front of a sensitive television camera (with a silicon-intensified target (SIT)), fitted with a 25 mm, $f/0.78$ lens with blue-green filter. The period of rotation of the retarder was 24 s. The polariser and retarder were Polaroid thin-sheet (cellulose acetate butyrate) and allowed wide field viewing; the SIT photocathode gave a field of view of $22^\circ \times 29^\circ$ horizontally. The noctilucent cloud images recorded on videotape can be analysed on playback using a video-integrator¹⁰ which allows the sum of video signals from any area selected in the television field of view to be displayed on a chart recorder. The amplitude of the signal is proportional to the average light intensity in the selected area.

Analysis of the video signal is best handled using Stoke's vector to describe the state of polarisation of light from the noctilucent clouds. The Stoke's vector consists of a set of four components (the Stoke's parameters I , M , C and S). Each parameter is a time-averaged intensity: I is the light intensity; M contains the 'horizontal preference' of the polarised part of the light, C the '45° preference' and S the 'circular preference'. These parameters may be written as a column vector and the transmission of polarised light through the Polaroid/retardation plate assembly is estimated using Mueller calculus¹¹. The state of polarisation of the light emerging from the Polaroid/retardation plate combination is found by successive multiplication of the Stoke's vector of the incident light by the two matrices representing the Polaroid and the retardation plate. The intensity I of light falling on the television camera varies with time such that

$$I(t) = X + Y \sin 4(\beta(t) + \beta_0) - Z \sin 2\beta(t)$$

where $\beta(t)$ is the instantaneous angle between the fast axis of the retarder and the (fixed) polariser axis. If the Stoke's parameters

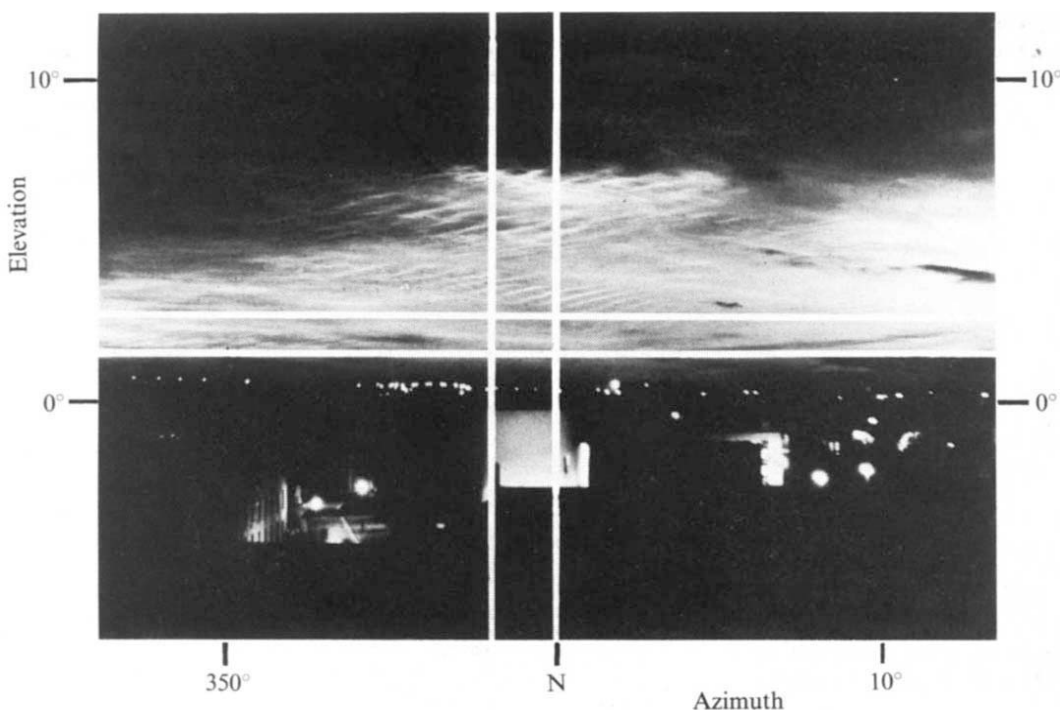


Fig. 1 Noctilucent cloud display seen from Aberdeen on the night of 28–29 July 1978. The print is from a monochrome copy of a 15-s ($f/2.8$) exposure on high-speed Ektachrome film. White lines enclose an area of sky where circular polarisation was detected.

of the light from the clouds are I_0 , M_0 , C_0 , S_0 , the expression for $I(t)$ contains the following:

$$\beta_0 = \frac{1}{4} \tan^{-1} (M_0/C_0);$$

$$2X = (k_1 + k_2)I_0 + \frac{1}{2}(k_1 - k_2)(1 + \cos \delta)M_0;$$

$$Y = (1/4)(k_1 - k_2)(1 - \cos \delta)(M_0^2 + C_0^2)^{1/2} \quad \text{and}$$

$$Z = \frac{1}{2}(k_1 - k_2) \sin \delta S_0$$

k_1 and k_2 are the principal transmittances of the polaroid; the retardance of the retarding plate is δ . In the simple case of partly polarised incident light which contains no circular polarisation ($S_0 = 0$), the video signal is a constant level X with a superimposed 4β sinusoidal modulation of amplitude Y . Addition of a small amount of circular polarisation ($S_0/I_0 \approx 0.01$) has the effect of adding in a 2β modulation on top of the 4β modulation and the presence of this shows up mainly as a half-frequency modulation Y . If the light from the clouds is predominantly linearly polarised, with only a small amount of circular polarisation added, then the detection of circular polarisation is simple, as extrema of the 4β signal will alternate in size.

Figure 1 shows a picture of the sky taken during the noctilucent cloud display of 28–29 July 1978. The Sun is 11° below the horizon at an azimuth of 24° . The videotape recording indicates that the degree of linear polarisation increases systematically from the clouds to the right of Fig. 1 (scattering angles $\leq 20^\circ$) to those shown in Fig. 1 (scattering angles 20 – 50°) as is to be expected both from scattering theory and from all previous reports of polarisation measurements of noctilucent clouds. However, regions in which the maxima of the 4β signal alternate in size have been found only in restricted parts of the cloud display and one such area is indicated in Fig. 1. Figure 2 shows part of the chart record obtained when integrating the video recording in this area. The alternation in height of the maxima is clearly seen, and analysis of the data from Fig. 2 shows the proportion of circular polarisation, S_0/I_0 , to be 0.005 ± 0.0013 . This is less than that of earlier measurements⁴, when S_0/I_0 was estimated to be typically 0.05 in the blue part of the visible spectrum. However, the noctilucent clouds that were seen on this occasion from Aberdeen were unremarkable in brightness and development, and the observation night was towards the end of the useful season.

The presence of a noticeable amount of circular polarisation in only certain parts of the display (adjacent regions showing, as far as can be seen, none) shows the effect is not produced locally by tropospheric scattering or absorption even at the great zenith distance of that area showing circular polarisation. There is no reason that the part of the noctilucent cloud that does show circular polarisation should have this difference from other parts of the display at the same zenith distance. It is also difficult to reconcile the appearance of localised circular polarisation with its being caused by partly polarised light incident on spherical particles. The cloud enclosed by the lines in Fig. 1 is at a latitude of 64.5° (no allowance being made for refraction) and there the Sun is only 4.4° below the horizon. The sunlight shining on the cloud passes the sunset line at a height of over 60 km and there is negligible absorption in the direct path. Any linear polarisation in the illumination of the cloud must come from tropospheric scattering near the sunset line, from that part of the atmosphere below the direct solar rays.

The relationship between degree of circular polarisation and amount of asphericity of the particles scattering the light cannot be described easily without knowing the distribution of sizes and shapes among the cloud particles. However, some estimate can be made by assuming that the low-contrast (or Rayleigh-Gans) approximations⁸ are applicable and that only unpolarised sunlight is shining on the clouds. If the scattering particles are cylinders of length exceeding two light wavelengths ($1 \mu\text{m}$), circular polarisation of $S/I = 0.005$ will arise if the radius of all the particles is close to $0.01 \mu\text{m}$. Alternatively, there could be a mixture of 10% of cylinders with radii $0.04 \mu\text{m}$ in a cloud of otherwise spherical scatterers.

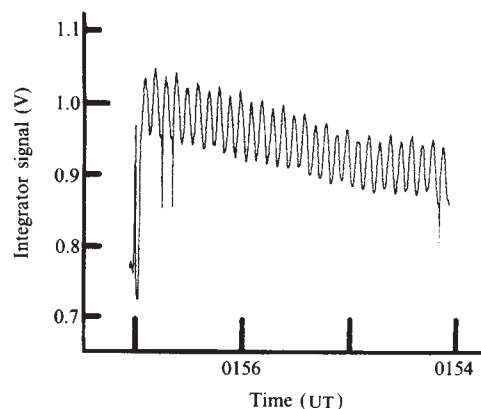


Fig. 2 Section of chart record made using a video-integrator on the area of sky indicated in Fig. 1. Note the depressed zero of the ordinate scale; the proportion of linear polarisation present is ~ 0.16 . The presence of a small amount of circular polarisation is shown by the alternation in height of the maxima.

Little is certain about the way in which ice crystals will grow in the low temperature and low pressure conditions near the mesopause. By analogy with the efficient screw-dislocation growth of crystals in the laboratory, it is possible that a significant proportion of the cloud particles will grow into needles rather than into spheres and this is certainly seen in the laboratory at low temperatures which are, however, much higher than the temperatures of noctilucent clouds. The fall speeds of needle-shaped crystals will be less than those of spheres. The amount of light scattered by a single particle increases very quickly as it grows in a supersaturated layer in the atmosphere. Spherical particles, however, will fall out of this layer more quickly than needle-shaped particles and so have less time for growing. It could be that there will be an aerodynamic advantage for growth of non-spherical particles relating to the spherical ones and that these non-spherical particles, being large, will individually scatter more light than the spheres and so seem to be relatively more important when observing the light scattered from a cloud.

The relative humidity of the atmosphere in the cloud region controls the optical thickness of the cloud that can form. Needle-like scatterers with their longer residence times and eventually greater relative sizes will give thicker (that is, brighter) clouds than spherical scatterers. For a given cloud brightness, the atmosphere does not have to be as wet with needles as with spheres.

M. GADSDEN

Department of Natural Philosophy,
Aberdeen University,
Aberdeen, UK

P. ROTHWELL
M. J. TAYLOR

Department of Physics,
Southampton University,
Southampton, UK

Received 25 January; accepted 13 March 1979.

1. Bronshten, V. A. & Grishin, N. I. *Noctilucent Clouds* (Jerusalem, 1976).
2. Smyth, C. P. *Nature* **34**, 311–312 (1886).
3. Witt, G., *J. geophys. Res.* **65**, 925–933 (1960).
4. Gadsden, M. *Ann. Geophys.* **33**, 363–366 (1977).
5. Tozer, W. F. & Beeson, D. E. *J. geophys. Res.* **79**, 5607–5612 (1974).
6. Donahue, T. M. *et al. J. atmos. Sci.* **29**, 1205–1209 (1972).
7. Willman, C. I. *et al. in Optical Investigations of the Emission of the Atmosphere, Aurorae and Noctilucent Clouds aboard the Orbital Scientific Station "Salyut-4"* (Tartu, 1977).
8. Banderman, L. W. & Kemp, J. C. *Mon. Not. R. astr. Soc.* **162**, 367–377 (1973).
9. Taylor, M. J. *J. sci. Instrum.* (in the press).
10. Crawford, J. thesis, Southampton Univ. (1979).
11. Shurcliff, W. A. *Polarised Light, Production and Use* (London, 1962).

Isotopically light carbon in diamonds from some kimberlite pipes in Lesotho

THIS study of the isotopic composition of carbon has been carried out on six diamond crystals from kimberlite pipes in Lesotho. One crystal came from the Kao pipe, one from the Liqhobong and four from the Kolo pipe. We consider here how their colour and location are related to their isotopic composition.

The Kao and Liqhobong kimberlite pipes are located in the highlands of the northern Lesotho where they cut through the thick basaltic lavas of the Drakensberg Beds^{1,2}. The Kolo pipe is situated in the lowlands of western Lesotho and it cuts through the shales and sandstones of the Beaufort series and the Karoo dolerites. Diamonds from these deposits are characterised by a large quantity of fragments and angular forms (up to 65% for the Kao pipe and 80% for the Liqhobong and Kolo pipes).

Generally among the crystals, rounded dodecahedra prevail^{3,4} but a tendency towards an increase of octahedral forms in the Liqhobong–Kao–Kolo direction is noticeable.

Table 1 A brief description of the diamonds and analytical results of isotopic composition of carbon

No.	Pipe	Brief description	¹³ C‰ PDB
1	Kao	A flattened rounded twin of dodecahedra, smoky, transparent fissured. Faces 110 are unevenly pitted. The lines of plastic deformation and traces of etching are observed. Weight 40.62 mg.	-4.6
2	Liqhobong	A fragment of a lemon-yellow transparent octahedron with triangular cavities and hillock of irregular or rounded shape on the relics of faces III. Weight 30.5 mg.	-20.5; -10.6
3	Kolo	An irregular intergrowth of 3 flattened colourless transparent octahedra with development of the curved dodecahedral surfaces. The surfaces are badly etched with cavities. The diamonds have several graphite inclusions. Weight 30.95 mg.	-16.2; -15.9
4	Kolo	A rounded badly elongated on L ₃ yellow-transparent dodecahedron with several cavities. Weight 30.95 mg.	-6.9
5	Kolo	A rounded flattened colourless dodecahedron with several graphite inclusions. Faces 110 are roughly pitted with some hillocks. Weight 30.81 mg.	-6.5
6	Kolo	A rounded flattened smoky dodecahedron. The faces are irregularly developed and have deep canals of etching. An inclusion of orange garnet is observed inside the crystal. Weight 50.83 mg.	-14.7; -15.1

Serial no. *	Crystal habit	Colour	¹³ C‰ PDB
1	Twin or rounded dodecahedron	Smoky	-4.6
2	Fragment of octahedron	Lemon-yellow	-20.5; -19.6
3	Intergrows of octahedra	Colourless	-16.2; -15.9
4	Rounded dodecahedron	Yellow	-6.9
5	Rounded dodecahedron	Colourless	-6.5
6	Rounded dodecahedron	Smoky	-14.7; -15.1

* The diamonds are described in the same succession as in the text.

Many of the stones are of different colours and varying intensity: for the Liqhobong pipe, yellow is dominant: Kolo diamonds are mostly smoky-grey whereas Kao diamonds are brownish. Most of the diamonds have black inclusions of graphite. A brief description of the diamonds which have been analysed is given in Table 1.

Analysis of the isotopic composition of carbon of the diamonds has been done by oxidation in a stream of oxygen on copper oxide. The relative contents of isotopes of carbon have been measured by a mass spectrometer (Varian MAT -230) with precision of $\pm 0.01\%$ ($\pm 0.1\%$). Analytical data in Table 1 are in values of ¹³C ($\pm 0.1\%$). Analytical data in Table 1 are in values of ¹³C (‰) which are deviations in ‰ of ¹³C/¹²C of samples with respect to ¹³C/¹²C of the PDB standard.

Table 1 shows that three of the six diamonds (nos. 2, 3 and 6) are characterised by somewhat light intermediate composition of carbon (group B of the Galimov *et al.*⁵ classification). Note that the analysis of these diamonds has been done twice and the data obtained have been rechecked.

Diamonds with a similar carbon isotope composition have not been reported either in kimberlite rocks of Africa^{6,7} or in Yakutia^{5,8}. Therefore, it had been thought until recently that the ratio ¹³C in kimberlites is rather constant and is characterised by a value ¹³C = -4 to -9‰ PDB the value assumed for mantle carbon carbonatites as also fall in this range⁹. The only exception were carbonados of Brazil with ¹³C = -27 to -30‰ PDB^{5,10} as well as some coloured diamonds from placers of the northern Yakutia¹¹. However, it has been established recently that among the diamonds from the placers of the Urals, Sayans, Ukraine and from other areas of the USSR there are monocrystals with light ¹³C = -20 to -25‰ PDB) and intermediate ¹³C = -10 to -20‰ PDB) isotopic composition of carbon⁵. Kaminsky *et al.*¹² have even suggested that such diamonds are of non-kimberlitic origin. In this connection, the discovery of isotopically light diamonds in Lesotho kimberlites is very significant. From these descriptions of the diamonds and from Table 1, neither the habitat of the stones nor their colours are related to a particular isotopic composition. Among the isotopically light diamonds there are rounded dodecahedra and octahedra which are colourless or have a yellow or smoky colour. This isotopic composition of diamonds probably depends on their genetic peculiarities but further study on this is required.

G. I. SMIRNOV
M. M. MOFOLO
P. M. LEROTHOLI

Department of Mines and Geology,
Maseru, Lesotho

F. V. KAMINSKY

Central Scientific Research,
Institute of Geological Prospecting,
Moscow, USSR

E. M. GALIMOV
I. N. IVANOVSKAYA

Institute of Geochemistry
and Analytical Chemistry,
A.N.U.S.S.R., Moscow, USSR

Received 15 November 1978; accepted 15 February 1979.

- Ralfe, D. G. *Lesotho Kimberlites* (ed. Nixon, P. H.) 101-105 (1973).
- Clement R. In *Lesotho Kimberlites* (ed. Nixon, P. H.), 110-121 (1973).
- Whitelock, T. K. In *Lesotho Kimberlites* (ed. by Nixon, P. H.), 128-140 (1973).
- Nixon, P. H. & Boyd, F. R. In *Lesotho Kimberlites* (ed. Nixon, P. H.), 141-148 (1973).
- Galimov, E. M., Kaminsky, F. V. & Ivanovskaya, I. N. *Geokhimiya* 340-349 (1978).
- Graig, H. *Geochim. cosmochim. Acta* 3, 53-92 (1953).
- Wickman, F. E. *Geochim. cosmochim. Acta* 9, 136-153 (1956).
- Vinogradov, A. P., Kropotova, O. I., Orlov, Y. L. & Grinenko, V. A. *Geokhimiya*, 1395-1397 (1966).
- Taylor, M. P., Frechen, J. & Degens, E. T. *Geochim. Cosmochim. Acta* 31, 407 (1967).
- Vinogradov, A. P., Kropotova, O. I. & Ystinov, V. I. *Geokhimiya* 643-657 (1965).
- Kovalsky, V. V., Galimov, E. M. & Prokhorov, V. S. *Doklady Acad. nauk. S.S.S.R.* 203, 440-442 (1972).
- Kaminsky, F. V., Galimov, E. M., Ivanovskaya, I. N., Kirikilita, S. I. & Polkanov, Y. A. *Doklady Acad. nauk S.S.S.R.* 236, 1207-1208 (1977).

New data on natural todorokites

TODOROKITE has been found as a product of terrestrial weathering, in sedimentary manganese ore, in deep-sea iron and manganese nodules and crusts. It is considered as an important source of Ni, Co, Zn, Cu, Mg in marine manganese nodules.¹⁻³

We show here that besides previously described todorokites with parameters $a_0 = 9.75$, $b_0 = 2.84$ and $c_0 = 0.6$ Å there are others, one with $a_0 = 14.6$ Å (The Pacific Ocean, Bakal) and one with $a_0 = 24.38$ Å (Bakal, Sterling Hill), b_0 and c_0 being unchanged. Morphologically the todorokites with different a_0 are very similar—their a_0 -values being multiples of 4.88.

According to Straczek *et al.*⁴, the approximate formula of todorokite might be $(\text{Na}, \text{Ca}, \text{K}, \text{Mn}^{2+})(\text{Mn}^{4+}, \text{Mn}^{2+}, \text{Mg})_6\text{O}_{12} \cdot 3\text{H}_2\text{O}$. The extent of variations of Mn^{4+} , Mn^{2+} , and Mn^{3+} are still unknown. The symmetry of the mineral is orthorhombic (or monoclinic, β about 90°), with $a_0 = 9.75$, $b_0 = 2.84$ and $c_0 = 9.59$ Å. In electron microscope preparations todorokite occurs in the form of thin lathes (001) frequently elongated along the b axis.

Burns and Burns^{1,5} assumed that the todorokite structure was similar to that of hollandite and psilomelane and had common features with the structure of rutile, this being shown by their ability to produce planar net trillings of elongated crystals. The intergrowths of todorokite form nets of hexagonal, triangular and rhombic cells. The electron micrographs of the treble intergrowths show three superimposed diffraction patterns orientated to each other with a relative rotation at an angle of 120° (Fig. 1). From selected area electron diffraction (SAD) todorokite has been shown to occur in the form of lathes with a basal face (001).

A study of SAD-patterns revealed that besides the todorokite having $a_0 = 9.75$, $b_0 = 2.84$ and $c_0 = 9.59$ Å found in many deposits, species with $a_0 = 14.6$ (Fig. 1c) and 24.38 Å (Fig. 1d), the b_0 and c_0 parameters being unchanged, occur in the natural environment. Under the microscope and on the electron micrographs these species look similar to todorokite with $a_0 = 9.74$ Å. The results, obtained from an analysis of individual lathes of todorokites with different values of the a_0 parameter carried out with an EM 100 C microscope and X-ray energy spectrometer (KeveX), indicate the similarity of their chemical composition. The X-ray powder patterns of these todorokite species display an insignificant difference between the parameters, which is reflected in slight variations of spacings for some reflections, hkl . The X-ray patterns failed to show weak reflections indicative of a true value of the a_0 parameter of todorokite. The strong reflection $hk0$ and 001 of structural varieties of todorokite largely correspond to the same lattice spacings.

It is easy to see that the values of a_0 characteristic for the members of the todorokite group (9.75, 14.6, 24.38 Å) are

divisible by 4.88 Å. The variations of this parameter are shown by the frequency of the reflections occurring along the axis a^* . For example on an electron micrograph showing the todorokite basal plane (001) with $a_0 = 9.75$ Å (Fig. 1b) there are three less-intense reflections along the a^* -direction between the strong reflections forming a pseudo-hexagonal net. The first strong reflection along the a^* axis has indices 400 ($d = 2.4$ Å). In the SAD pattern of the todorokite with $a_0 = 14.6$ Å (Fig. 1c) there are five weak reflections between the strong ones. In this case, the reflection at $d = 2.44$ Å along the a axis has indices 600. On the pattern of todorokite with $a_0 = 24.40$ Å (Fig. 1d) there are, between the strong reflections along the a axis, four weak reflections on each side, which, provided other reflections are absent, would determine a base-centred unit cell and the extinction law for the plane $(001)^* - h + k = 2n$, where n is an integer. However, on the diffraction pattern of the todorokite species with $a_0 = 24.38$ Å, midway between the strong reflections, there is the fifth, weak, reflection which violates the symmetry of the unit cell and shows it to be really primitive. In this case, the strong reflection at $d = 2.44$ Å occurring on the a^* axis has indices 10.0.0.

Under the microscope todorokite with the parameter $a_0 = 14.6$ Å is seen to occur in the nucleus of an iron-manganese nodule collected from the Pacific Ocean floor. Flaky todorokite from the Bakal deposit (USSR) seems to consist of aggregates of two todorokite species with the a parameters 14.6 and 24.4 Å. Todorokite with $a_0 = 25$ Å has been identified in a sample from Sterling Hill (USA) and a specimen from the Takhta-Karacha deposit in the Middle Asia.

Thus, todorokite species may be characterised by unit cells having different parameters a_0 which is divisible by 4.88 Å (the parameter along the a axis being determined by strong reflections located in accordance with the pseudo-hexagonal law, if there is a base-centred orthogonal unit cell). We assume that todorokites with other a_0 parameters, but still divisible by 4.88 Å, do also exist in nature.

The variations of the a parameter and formation of the todorokite superstructures will probably be attributed to certain difference in their composition and the environment of their crystallisation.

The data available show that todorokite is displaced by vernadite (δMnO_2). Even after a complete displacement of todorokite by vernadite, the former retains its fibrous or flaky appearance. Therefore, the products of partial or complete replacement of todorokite by vernadite can be erroneously taken for todorokite, without comprehensive electron microscopic examination. Partially vernadite-replaced todorokite as well as the pseudomorphs of vernadite after todorokite are common.

In view of the possibility of a new species of todorokite family being found in nature as well as in synthetic products, we recommend the following nomenclature: todorokite I with $a_0 = 4.88$ Å; todorokite II with $a_0 = 4.88$ Å $\times 2$; todorokite III with

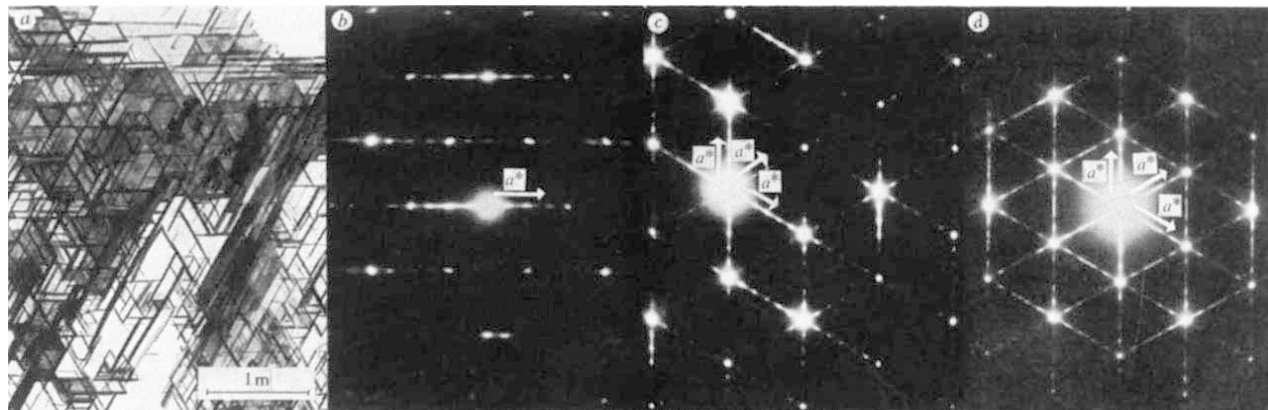


Fig. 1 *a*, electron photomicrographs of the todorokite treble intergrowths. *b*, *c*, *d*, Electron diffraction pattern of monocrystal of todorokite with parameter $a_0 = 9.75$ Å (*b*), and of treble intergrowths of todorokite with parameters $a_0 = 14.6$ Å (*c*) and 24.4 Å (*d*).

$a_0 = 4.88 \text{ \AA} \times 3$; todorokite IV with $a_0 = 4.88 \text{ \AA} \times 4$; todorokite V with $a_0 = 4.88 \text{ \AA} \times 5$. (Todorokites II, III and V have been already found in nature.)

Predictions made by other authors have a direct bearing on the discovery of todorokite polymorphs. They deal with pure crystalline manganese oxides and hydroxides (see refs 5, 6). According to Turner and Buseck⁷ mixtures of polymorphs found recently have also been known for other manganese oxides—hollandites and cryptomelanes.

F. V. CHUKHROV
A. I. GORSHKOV
A. V. SIVTSOV
V. V. BERESOVSKAYA

Institute of Ore Geology,
Staromonety 35,
Moscow, USSR

Received 24 October 1978; accepted 12 February 1979.

1. Burns, R. G. & Burns V. M. *Phil. Trans. R. Soc. A* **285**, 249–258 (1977).
2. Burns, V. M. & Burns, R. G. *Am. Miner.* **63**, 827–831 (1978).
3. Burns, V. M. & Burns, R. G. in *Sur La Genèse des Nodules de Manganese* (éd. Laloe, C.) (Colloq. Int. CNRS, no. 289, in the press).
4. Straczek, J. A., Horen, A., Ross, M. & Warsaw, Ch. M. *Am. Miner.* **45**, 1174–1184 (1960).
5. Burns, R. G. & Burns, V. M. in *Marine Manganese Deposits* (éd. Glasby, G. P.) 185–248 (Elsevier, Amsterdam, 1977).
6. Wadsley, A. D. in *Non-Stoichiometric Compounds* (éd. Mandelcorn, L.) 98–209 (Academic, New York, 1963).
7. Turner, S. & Buseck, P. R. *A. Meet. geol. Soc. Am. Abstr.* **10**, 507 (1978).

Rates of Holocene folding in the coastal Zagros near Bandar Abbas, Iran

FOR the study of recent crustal processes, seismological and geodetic surveys need to be complemented by measurements of Earth movements over periods long enough for secular trends to emerge¹. The one such item available for the Zagros was obtained in 1952, when Lees and Falcon used archaeological remains to trace the growth over the past 1,700 yr of the Shaur anticline near the head of the Persian Gulf². Lees and Falcon also observed that recent marine terraces on the shores of the Gulf had been deformed by anticlinal growth²: I report here

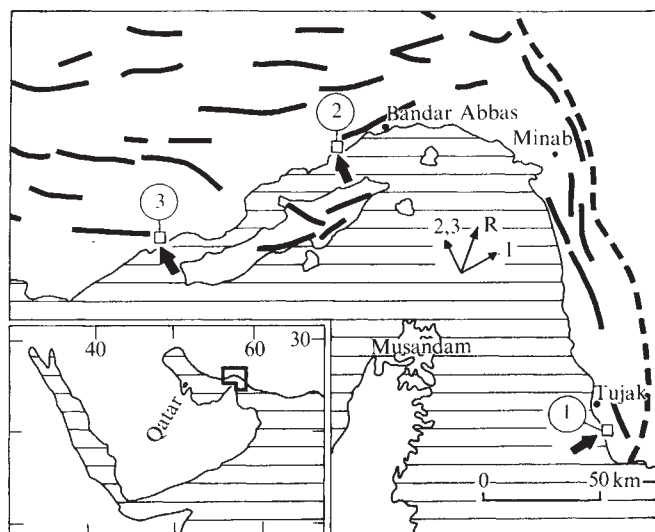


Fig. 1 Location of sections. Main fold axes (heavy lines) after Berberian¹⁷. Broken heavy line shows approximate position of Minab Fault. Arrows are normal to generalised strike of corresponding anticline and denote shortening vectors; bearing of resultant **R** (Table 2) based on vectors at sections 1 and 2, which are at right angles.

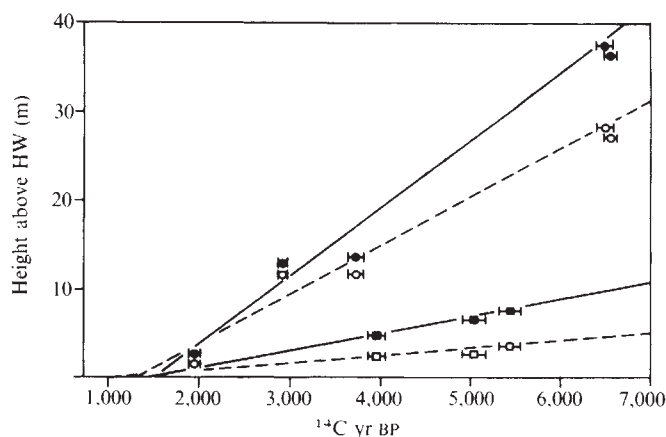


Fig. 2 Age/height plots for section 1 (●) and sections 2 and 3 (■) corrected for eustatic factor after Flint⁷; corresponding open symbols (○, □) corrected after Clark *et al.*⁸. Solid lines are least-squares fits for former, dashed lines for latter. Age errors bars denote 1 s.d. No correction has been made for 'apparent age' at death of molluscs, which would shift all points to the left.

preliminary ¹⁴C dates on shells from fossil beach deposits associated with three fold structures on the eastern Zagros coast of Iran near Bandar Abbas, and derive Holocene deformation rates from them. The results, though not necessarily applicable to other parts of the folded belt, are compatible with estimates of the current rate of rotation of Arabia, to whose northeasterly drive the Zagros Range has been ascribed³.

The samples were collected for a survey of recent coastal movements⁴ which, in view of the uncertainty that surrounds global (eustatic) sea-level history, is confined to the relative displacement of isochronous shorelines⁵. For the present work, however, absolute values of vertical displacement are needed and hence allowance must be made for fluctuations in sea level. Data from other parts of the Gulf cannot be used for this as they may embody the tectonic effects we are trying to isolate. For this reason, and also to facilitate eventual comparison with data from the Makran⁶, the corrections adopted below are based on the empirical global curves favoured by Flint⁷, which indicate progressive if decelerating submergence during the Holocene, and on the model formulated by Clark *et al.*⁸, which for the latitudes at issue yields slow emergence over the past 7,500 yr.

The folds of the southern Zagros east of Bandar Abbas swing round parallel to the coast. The first section considered here, 2 km south of Tujak (1 on Fig. 1), consists of a succession of beaches cut into marls of the Mio-Pliocene Agha Jari Formation which were exposed by erosion of an anticline striking N 160° E. The beaches are capped by 0.5–1.5 m of fossiliferous sands whose lithology, form and faunal content reflect intertidal conditions⁹. Three of the beaches were sampled for dating. Each sample consisted of a single species of mollusc; thin-section analysis, X-ray diffraction and inspection under the scanning electron microscope were used to identify unaltered specimens.

The second set of dates refers to two anticlines west of Bandar Abbas which are also composed of late Tertiary rocks. Unrecrystallised shells were obtained from the summit of a beachrock on the southern flank of one of the anticlines, which strikes roughly N 60° E and consists of sands and marls of the Miocene Mishan Formation (3 on Fig. 1), and from an abandoned offshore bar and a beach terrace on the southern flank of the other, which strikes N 70° E and includes both Agha Jari and Mishan deposits (2 on Fig. 1). The results from sections 2 and 3 are treated together; additional dates should make it possible to establish differences between the histories of the two structures.

The ¹⁴C ages obtained for both coasts are listed in Table 1 and, after eustatic correction, plotted against height in Fig. 2. The presence of successive discrete beaches, notably at section 1,

Table 1 Radiocarbon ages for coasts east and west of Bandar Abbas

Sample no.	Laboratory no.	Location	Age (yr BP)*	$\delta^{13}\text{C}(\text{‰})$	Species	Elevation above HW
Section 1						
ET Ia	SRR 1258	26°00' N, 57°14' E	1,940 ± 45	+0.40	<i>Scapharca inaequivalvis</i>	2.0
ET IIb	UM 1268	"	3,735 ± 85	+0.79	<i>S. inaequivalvis</i>	11.8
ET IIc	SRR 1257	"	2,925 ± 40	+1.90	<i>Asaphis deflorata</i>	11.8
ET IIIb	UM 1269	"	6,525 ± 95	+0.46	<i>S. inaequivalvis</i>	28.6
ET IIIx	UM 1255	"	6,580 ± 155	+0.83	<i>A. deflorata</i>	28.6
Section 2						
76/1	MC 1574	27°09' N, 56°08' E	3,960 ± 100	+2.86	<i>Turritella fultoni</i>	2.3
76/4a	MC 1572	"	5,410 ± 110	+2.96	<i>T. fultoni</i>	4.0
Section 3						
IBR/1	MC 1099	26°47' N, 55°10' E	5,040 ± 115	+1.42	<i>Barbatia cf. fusca</i> and <i>A. deflorata</i>	3.0

* Corrected for isotopic fractionation by normalising to a $\delta^{13}\text{C}$ value of -25‰ relative to PDB.

suggests that movement was spasmodic, but average uplift rates can still be informative. Table 2 lists the values obtained both by assuming uniform uplift since the formation of the highest sea-level indicator and by calculating least-squares lines for the data. The maxima obtained are 7.4 mm yr^{-1} for section 1 and 1.9 mm yr^{-1} for sections 2 and 3. These figures lie between the minimum average of 1 mm yr^{-1} that has been proposed for general uplift of the Zagros Mountains³ since the early Pliocene, and the rate of about 12 mm yr^{-1} obtained for the Shaur anticline². A Holocene rate of 6 mm yr^{-1} has been calculated for subsidence of the Musandam Peninsula at the mouth of the Gulf¹⁰; data from Qatar¹¹ give uplift rates of 1.1 and 0.3 mm yr^{-1} according to which eustatic correction is adopted.

A preliminary assessment of the horizontal shortening represented by the observed uplift can be made using purely geometrical methods and assuming that the folds approximate

The values for ΔX for sections 2 and 3, which represent northward compression, are probably minima, as shortening is likely to be taken up in part by other anticlines inland and offshore. Folding due to eastward shortening at section 1 is confined to a narrow belt by the Minab Fault to the east¹², but evidence at the Chah Shirin Fault near Minab for 3.5 m of north-eastward thrusting over the past 7,000 yr (ref. 15) suggests that on this coast also the calculated shortening is a minimum. At all events the ΔX values give results (Table 2) which range between 16.6 and 31.9 mm yr^{-1} and bear roughly NNE. Palaeomagnetic evidence from the Red Sea and the Gulf of Aden indicates separation between Africa and Arabia at a rate of about 20 mm yr^{-1} during the past 5 Myr and a spreading axis in the western Gulf of Aden with a strike of $\text{N } 110^\circ \text{ E}$ (ref. 16). The relationship between the two sets of results needs to be scrutinised.

Table 2 Uplift and compression rates

	$L (\times 10^5 \text{ mm})$	$s (\times 10^5 \text{ mm})$	ω (rads)	Average (mm yr^{-1})				Least-squares plot (mm yr^{-1})			
				a		b		a		b	
				Δy	ΔX	Δy	ΔX	Δy	ΔX	Δy	ΔX
Section 1	245.0	61.25	1.3788	5.7	20.0	4.3	15.1	7.4	26.0	5.4	19.0
Sections 2 and 3	167.5	15.0	0.4887	1.4	13.6	0.7	6.8	1.9	18.4	0.8	7.8
Resultant (mm yr^{-1})				24.2		16.6		31.9		20.5	
Bearing				N 26° E		N 36° E		N 25° E		N 38° E	

Heights corrected for eustatic factor after Flint⁷ (a) and Clark *et al.*⁸ (b).

symmetrical curves of intrinsic sine form. Let L be the length of the fold surface between two troughs and ω the maximum slope of the limbs. During compression with horizontal displacement ΔX , the vertical displacement Δy of a point P on the fold at a distance s along the fold limb from one of the troughs can be approximated (C. D. Mann, personal communication) by the expression

$$\Delta y = \Delta X \times 2\pi/\omega \times (s/L)^2$$

where s/L is small and the folded zone is bounded at the base by a surface of discontinuity (décollement)^{3,12}. Known values for L , ω , s and Δy , and calculated values of ΔX , are given in Table 2. L , s and ω were derived from field measurements and air photographs for sections 1 and 3, and from geological cross-sections constructed from published maps¹³ by the Busk method¹⁴ for section 2. The many minor faults associated with all three structures have been disregarded.

Movement on the Chah Shirin Fault had apparently ceased by 1,250 yr BP¹⁵; the least-squares plots for section 1, like those for sections 2 and 3, suggest that uplift stopped 1,300–1,000 yr BP. Further age determinations and field mapping are required to establish whether this is an artefact of dating and to refine the uplift plots so that the underlying mechanisms can be explored.

I thank M. G. Ghorashi, C. D. Mann, D. J. Shearman, J. D. Taylor and B. M. Walker for help, the Central Research Fund of London University, the Centre Scientifique de Monaco, the NERC and the Royal Society for radiocarbon dates and isotopic data, the Geological Survey of Iran for field support, and R. H. Sibson for helpful comments.

C. VITA-FINZI

University College,
Gower Street,
London WC1, UK

Received 15 January; accepted 27 February 1979.

1. Pavoni, M. *Bull. R. Soc. N.Z.* **9**, 7–17 (1971).
2. Lees, G. M. & Falcon, N. L. *Geogr. J.* **118**, 24–39 (1952).
3. Falcon, N. L. in *Mesozoic-Cenozoic Orogenic Belts* (ed. Spencer, A. M.) 199–211 (Scottish Academic Press, Edinburgh, 1974).
4. Vita-Finzi, C. *Geogr. J.* **141**, 415–420 (1975).
5. Vita-Finzi, C. *Recent Earth History* (Macmillan, London, 1973).
6. Page, W. D., Alt, J. N., Cluff, L. S. & Plafker, G. *Tectonophysics* **52**, 533–547 (1979).
7. Flint, R. F. *Glacial and Quaternary Geology* (Wiley, New York, 1971).
8. Clark, J. A., Farrell, W. E. & Peltier, W. R. *Quat. Res.* **9**, 265–287 (1978).
9. Vita-Finzi, C. *Bull. Geol. Surv. Iran* (in the press).
10. Vita-Finzi, C. *Geogr. J.* **139**, 414–421 (1973).
11. Vita-Finzi, C. in *Qatar Archaeological Report. Excavations 1973* (ed de Cardi, B.) 11–25 (Oxford University Press, 1978).
12. Shearman, D. J. *Geogr. J.* **142**, 393–410 (1976).
13. Geological Map of S.W. Iran, 1/250,000 (British Petroleum Company, London, 1963).
14. Busk, H. G. *Earth Flexures* (Cambridge University Press, 1929).
15. Vita-Finzi, C. & Ghorashi, M. *Tectonophysics* **44**, 21–25 (1978).
16. Girdler, R. W. & Styles, P. *Nature* **271**, 615–617 (1978).
17. Berberian, M. *Seismotectonic Map of Iran*, 1/250,000 (Geological Survey of Iran, Tehran, 1976).

Ordovician fish spines from Girvan, Scotland

THE Upper Ordovician Lady Burn Starfish Beds¹ of the Drummuck Group² in the Craighead inlier have long been known for their well-preserved, diverse and abundant invertebrate faunas^{3,4}. The trilobite⁵ and brachiopod (D.A.T.H., unpublished data) faunas indicate a late Rawtheyan (Ashgill) age for these predominantly green medium grained sandstone beds. Moreover, a study of the brachiopod associations and sediments of these beds together with the rocks above and below suggests that, although the Starfish Beds contain essentially 'shelf' faunas, final deposition was in a relatively deep-water, proximal slope environment⁶. I present here geochemical and morphological evidence of fish spines from these deep-water, siliciclastic marine rocks of Ordovician age at Girvan, south-west Scotland.

The occurrence of a pair of fish spines in the oldest and most fossiliferous of these marine strata at the South Threave locality (GR NS 250 038) is an unusually early appearance of this type of vertebrate material in the fossil record. The spines (Fig. 1) are subparallel and lie on a very fossiliferous bedding plane, which is crowded by disarticulated and randomly orientated valves of the Ashgill brachiopod *Eochonetes advena* Reed, 1917. The better preserved spine is 13 cm long, broken proximally and slightly curved. It is composed of predominantly black, apparently fibrous and lustrous material. Although now somewhat flattened, the spine seems to have been crudely rectangular in cross-section and has a hollow central canal.



Fig. 1 Ordovician fish spines embedded in fossiliferous bedding plane.

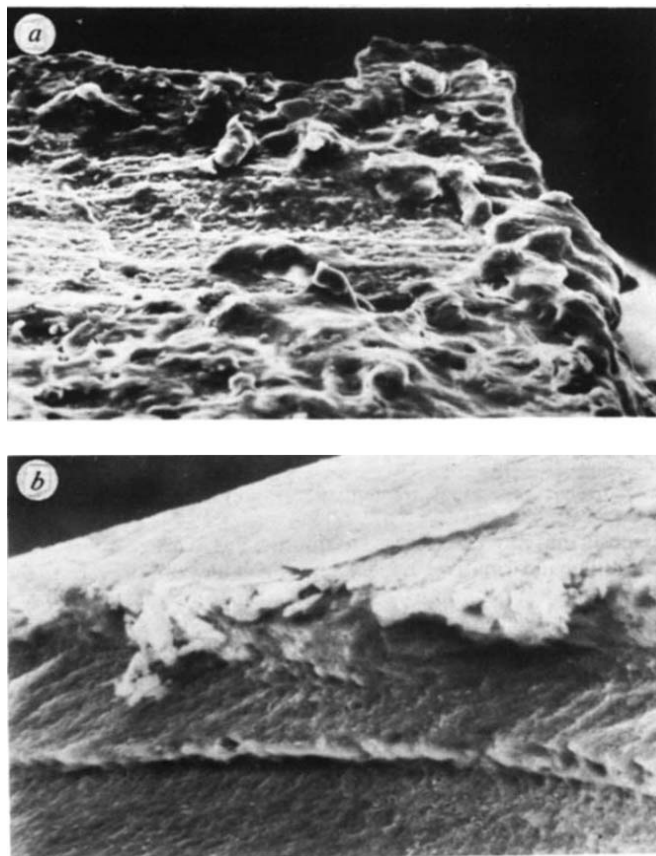


Fig. 2 SEM photomicrographs of spine fragment showing relatively simple bone histology. a, Oblique transverse view; b, oblique longitudinal view.

Electron microprobe analysis (D. C. Johnston, personal communication) indicates the composition of the material to be close to that of francolite, a carbonate-hydroxyfluorapatite commonly found as vertebrate bone.

Scanning electron microscope (SEM) photomicrographs (Fig. 2) of fragments of the material show the bone histology to be relatively simple; the oblique transverse (Fig. 2a) and longitudinal (Fig. 2b) views demonstrate the microstructure to be fairly massive, but with well developed growth rings about 20 μm apart. Interpretation of the cellular structure is difficult but in transverse section numerous minute pores perforate the substance; these presumably link with the central canal.

The morphological and geochemical evidence thus clearly demonstrate that these fossils are composed of vertebrate bone material and are best interpreted as fish spines; unfortunately no fish scales have been recovered.

Fish material is rare in rocks of pre-Silurian age, and when such remains are found they tend to be fragmentary⁷. The Ordovician occurrences, for example the Lower Ordovician Valhalla Formation of Spitzbergen⁸ and the Middle Ordovician Harding Formation of Colorado⁹, are usually represented by disarticulated heterostracan scales and cartilage. Similarly the earlier Cambrian record¹⁰ is marked by heterostracan plate fragments and tubercles. Although the similarly fragmentary nature of the later Girvan material does not allow an unequivocal determination, it cannot at present be related to any known fossil of Ordovician age. Most probably the spines are those of an acanthodian (spiny shark); note that the acanthodians are particularly characterised by paired fin spines¹¹. The close association of the spines suggests that the fish, or at least that part of the fish bearing the spines, had undergone negligible transportation since falling on to the substrate.

Recent studies of Palaeozoic fish¹² record the earliest appearance of acanthodians as Upper Silurian; in rocks of this age these fish are again represented by disarticulated spines and scales. The evidence from the Drummuck material, which strongly suggests that the pair of spines was part of a large specimen of acanthodian, significantly extends the stratigraphical range of these fish back into the Upper Ordovician.

The small amounts of the present material and lack of any comparable pre-Silurian occurrences do not allow much discussion on the origins of these spiny fish. It is, however, probable that by the later Ordovician the first marine vertebrates included at least both acanthodian and heterostracan fish.

I thank R. E. H. Reid for his encouragement and Professors D. L. Dineley and A. D. Wright for their comments. R. Reed and his staff assisted with SEM work and the plates are by Seán Watters.

D. A. T. HARPER

Department of Geology,
Queen's University,
Belfast, Northern Ireland

Received 23 January; accepted 9 March 1979.

1. Begg, J. L. *Trans. geol. Soc. Glasg.* **21**, 29–47 (1946).
2. Lamont, A. *Trans. geol. Soc. Glasg.* **19**, 288–334 (1935).
3. Lapworth, C. Q. *Jl geol. Soc. Lond.* **38**, 537–664 (1882).
4. Peach, B. N. & Horne, J. *The Silurian Rocks of Britain. 1, Scotland. Mem. geol. Surv. U.K.* (1899).
5. Ingham, J. K. *Proc. York. geol. Soc.* **35**, 455–505 (1966).
6. Harper, D. A. T. in *The Caledonides of the British Isles* (in preparation).
7. Winder, C. G. in *The Ordovician System* (ed. Bassett, M. G.) 645–657 (Cardiff, 1976).
8. Bockelie, T. & Fortey, R. A. *Nature* **260**, 36–38 (1976).
9. Spjeldnaes, N. *Colloques int. Cent. natn. Rech. scient.* **163**, 11–20 (1967).
10. Repetski, J. E. *Science* **200**, 529–531 (1978).
11. Watson, D. M. S. *Phil. Trans. R. Soc.* **228**, 49–146 (1937).
12. Moy-Thomas, J. A. & Miles, R. S. *Palaeozoic Fishes* (London, 1971).

Growth rings in dinosaur teeth

ANNUAL growth rings are often found in the skeletal tissues of Recent and fossil ectothermic vertebrates^{1,2}. These rings are usually attributed to the inability of ectotherms to maintain high levels of activity, feeding and growth during adverse times, such as dry seasons in tropical areas and cold seasons at higher, more temperate latitudes¹. In endothermic vertebrates, however, the efficient production of endogenous heat results in a relatively constant body temperature, and high activity and feeding levels are more easily sustained. Consequently, the growth of skeletal tissues in these animals is less sensitive to seasonal variations than in ectotherms, and annulations are rare except in species living under severe seasonal regimes^{3,4}. Many workers have proposed that dinosaurs were endothermic, as are all living mammals and birds, rather than ectothermic, as are all living reptiles^{3,5–10}. In accordance with this hypothesis, pronounced seasonal growth rings have not been observed in the calcified tissues of dinosaurs^{5,6}. However, here I report the discovery of such rings in the teeth of diverse Late Cretaceous dinosaurs.

Underwater screening for small vertebrate fossils has led to the recovery of hundreds of dinosaur teeth from the non-marine Milk River (early Campanian), Oldman (middle Campanian) and Edmonton (Maestrichtian) formations of Alberta, Canada. The teeth include those of Saurischia (Tyrannosauridae (Fig. 1) and Coeluridae) and Ornithischia (Hadrosauridae (Fig. 2a), Ceratopsidae (Fig. 2b) and Ankylosauridae). Ground and polished sections, as well as acetate peels of etched surfaces, have revealed conspicuous rings in the dentine in all dinosaur groups examined. Two main types of rings occur (Fig. 1): first, contour lines of Owen, which in Recent species mark minor nutritional and metabolic variations during the deposition of

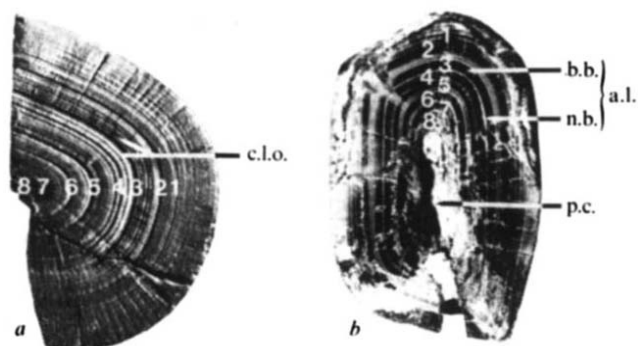


Fig. 1 *a*, Cross-section of a tyrannosaurid tooth, Milk River Formation, Alberta; c.l.o., contour lines of Owen; 1–8, annual layers in order of deposition ($\times 7.2$). *b*, Cross-section of a tyrannosaurid tooth, Oldman Formation, Alberta; n.b., narrow band; b.b., broad band; a.l., annual layer; 1–8, annual layers in order of deposition; p.c., pulp cavity ($\times 4.2$).

dentine and are accentuated according to the degree of disturbance experienced¹¹; and second, larger rings matching those of various Recent species^{4,12}, in which a narrow hypercalciified band (winter's growth) and a broad hypocalcified band (summer's growth), produced by seasonal variation in growth rate of the organic stroma of dentine, constitute an annual layer¹². Histologically, then, the structure of the dentine of dinosaurs indicates a periodicity in growth; the significance of this for the controversy concerning dinosaurian physiology is apparent when the clarity and pattern of the rings are compared to those in Recent and fossil organisms of known thermal physiology.

The only living terrestrial endotherms with annual layers of dentine as evident as those in Albian dinosaurs are mammals of Arctic or temperate regions^{4,12,13} subject to marked seasonal contrasts. Of tropical and subtropical mammals, only certain Australian marsupials^{4,12} and various African ungulates^{14–17} have been extensively investigated and no distinct annual layers in the dentine or bone have been reported, although layering in the cementum of some of the ungulates may correlate with wet and dry seasons. Consequently, if dinosaurs were endothermic and comparable to mammals in their thermal physiology, seasonal contrasts in Alberta during the Late Cretaceous must have been comparable to those of temperate and Arctic North America and Eurasia today to have produced conspicuous annual layers in the dentine.

However, severe seasonal contrasts do not seem to have characterised Alberta during the Late Cretaceous: palaeobotanical¹⁸ and palaeontological studies¹⁹ indicate a warm, subtropical climate without extreme daily or seasonal temperature change throughout the western interior of North America. If dinosaurs were endothermic in these conditions, how are the pronounced growth rings in their dentine to be explained? The



Fig. 2 *a*, Cross-section of a hadrosaurid tooth, Oldman Formation, Alberta, Canada ($\times 3.15$). *b*, Cross-section of a ceratopsid tooth, Oldman Formation, Alberta, Canada ($\times 5.04$).

origin of the rings is clearer, however, if dinosaurs were ectothermic, and thus presumably sensitive to relatively mild seasonal contrasts of temperature.

Fossil crocodile teeth (Fig. 3a) from the same localities as the dinosaur teeth have well developed rings, particularly similar to those of tyrannosaurids. No appreciable difference in pattern or clarity of annual layers or contour lines of Owen was found between the dentine of crocodiles (which are known ectotherms) and that of dinosaurs, suggesting that these organisms had a similar physiological tolerance to environmental fluctuations and hence a similar thermal physiology. Crocodile teeth from different geological settings, from Palaeocene and Eocene beds in Saskatchewan and Wyoming, also have growth rings like those of the Upper Cretaceous crocodiles and dinosaurs, thereby eliminating preservational bias.

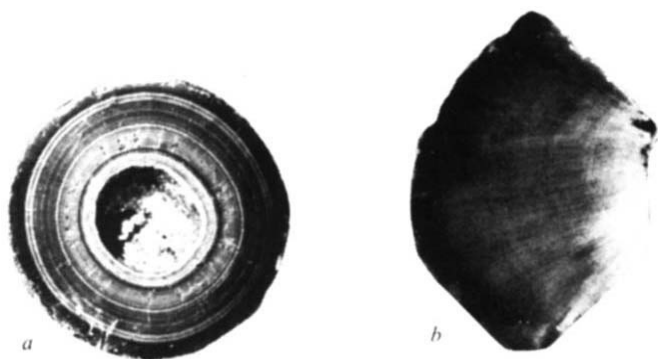


Fig. 3 a, Cross-section of a crocodile tooth (probably *Leidyosuchus* sp.), Oldman Formation, Alberta, Canada ($\times 3.6$). b, Cross-section of the root of an incisor of *Entelodon* sp., Oligocene, Cypress Hills Formation, Saskatchewan, Canada ($\times 3.0$).

Undoubted endotherms of large body size had not evolved by the Late Cretaceous and thus cannot be compared with dinosaurs. In teeth of small Late Cretaceous endotherms (mammals such as *Protungulatum*, *Didelphodon* and *Gypsonictops*) dentinal rings were not observed. Teeth from Tertiary mammals (representing four orders—Condylarthra, Amblypoda, Perissodactyla and Artiodactyla—and from formations interpreted as representing subtropical to warm temperature regimes) did show contour lines of Owen (Fig. 3b); however, apparent annual banding was rare and poorly developed.

As annual layers in most dinosaur and crocodile teeth are appreciably more conspicuous than in the Tertiary mammalian specimens, these reptiles were evidently affected to a greater degree by environmental factors influencing their growth rates. Furthermore, in the dinosaur and crocodile teeth, but not in the Tertiary mammal teeth, contour lines of Owen were often accentuated and tightly packed near and within the narrow bands of the annual layers: perhaps the crocodiles and dinosaurs were more sensitive to minor metabolic and nutritional stresses, particularly near and during adverse seasons, than were the Tertiary mammals examined.

In summary, the growth rings in the dentine of dinosaurs are more like those of contemporary crocodiles (known ectotherms) than those of fossil or Recent mammals from comparable climatic regimes. Evidently, growth and calcification of dentine in Late Cretaceous dinosaurs and crocodiles were affected similarly by both short-term and seasonal environmental variations, suggesting that these animals had similar thermal physiologies. If dinosaurs were ectothermic, their annual layers (as well as those of crocodiles and other organisms such as *Champsosaurus*, *Lepisosteus* and *Taxodioxylon* (unpublished data from the Late Cretaceous of Alberta)) seem best explained by Dodson's hypothesis²⁰ of seasonally distributed rainfall; by analogy, Recent crocodiles in Ngamiland exhibit annual layers in the periosteal bone apparently due to temperature fluctuations associated with wet and dry seasons characteristic of that tropical area¹.

R. C. Fox provided funds for this study (National Research Council of Canada grant A-3408) and I thank him for reading the manuscript.

PAUL A. JOHNSTON

Department of Geology,
University of Alberta,
Edmonton, Alberta,
Canada T6G 2E1

Received 15 August 1978; accepted 26 February 1979.

1. Peabody, F. E. *J. Morph.* **108**, 11 (1961).
2. Warren, J. W. *Diss. Abstr.* **24**, 908 (1963).
3. Bakker, R. T. *Scient. Am.* **232**, 58 (1975).
4. Morris, P. *Mam. Rev.* **2**, 69 (1972).
5. Ricqlès, A. *de Evol. Theory* **1**:51 (1974).
6. Ricqlès, A. *de in Morphology and Biology of Reptiles* No. 3 (eds Bellairs, A. d'A. & Cox, C. V.) 123 (Linn. Soc. Symp. Ser., 1976).
7. Bakker, R. T. *Evolution* **25**, 636 (1971).
8. Bakker, R. T. *Nature* **238**, 81 (1972).
9. Ostrom, J. H. *Evolution* **28**, 491 (1974).
10. Dodson, P. *Evolution* **28**, 494 (1974).
11. Schour, I. & Hoffman, M. M. *J. dent. Res.* **18**, 91 (1939).
12. Klevezal', G. A. & Kleinenberg, S. E. (1967), *Age Determination of Mammals from Annual Layers in Teeth and Bones* (in Russian) (Translation by Israel Program for Scientific Translations, Jerusalem, 1969).
13. Morris, P. in *Development Function and Evolution of Teeth* (eds Butler, P. M. & Joysey, K. A.) 483 (Academic, London, 1978).
14. Spinage, C. A. E. *Afr. Wildl. J.* **5**, 1 (1967).
15. Spinage, C. A. E. *Afr. Wildl. J.* **11**, 165 (1973).
16. Simpson, C. D. & Elder, W. H. *Arnoldia* **4**, 1 (1969).
17. Robinette, W. L. & Archer, A. L. E. *Afr. Wildl. J.* **9**, 83 (1971).
18. Axelrod, D. I. & Bailey, H. P. *Evolution* **22**, 595 (1968).
19. Estes, R. *Univ. Calif. Publ. Geol. Sci.* **49**, 1 (1964).
20. Dodson, P. *Palaeogeog. Palaeoclimatol. Palaeoecol.* **10**, 10 (1971).

Analogous motion illusion in man and fly

SINCE Wertheimer's classic paper¹, research in motion perception has been concerned with the study of visual illusions such as ϕ -motion. Various phenomena of this type are easy to elicit by successive changes of the light flux in spatially distinct photoreceptors, and easy to explain by the specific properties of the motion detectors, although there are reports to the contrary². The present account deals with a less easily comprehensible illusion which is elicited by simultaneous changes of the light flux in differently illuminated receptors³⁻⁵. The phenomenon has previously been ascribed to the prolonged latency of the 'light-on' responses at lower levels of illumination which converts simultaneous stimuli into successive signals⁶, but this does not explain the illusion satisfactorily³. MacKay and co-workers were the first to attribute the apparent motion to the adaptive properties of the input channels of the motion detectors⁴. We show that the illusion can be induced in the fruitfly, *Drosophila melanogaster*, as well as in man. The course control response to motion provides a quantitative assay of the illusion in the fly. The results suggest that the illusion originates in the distortion of the visual signals before motion detection.

To elicit the illusion in humans we illuminated transparent optical wedge filters of linearly graded density with a variable light source. The observer of the circular array in Fig. 1a usually perceives a spurious 'propeller' behind the pattern which rotates a few degrees towards the less transparent side of the wedge filters during light increase, and in the reverse direction during light decrease. The predominant sensation is oscillatory rotation at modulation frequencies of 0.3–2 Hz, continuous rotation towards the less transparent side of the filters at 5–12 Hz and ambiguity at 2–5 Hz.

Motion around a fly evokes vigorous turning in the direction of the horizontal displacement. The response counteracts involuntary deviations from a straight course in a stationary environment. Turning is induced by a regular network of elementary motion detectors. The network accepts stimuli from all parts of the visual field, and acts on the difference between the

propulsive forces on either side. With the fly held in a stationary position the course control responses of wings⁷ or legs⁸ can be measured quantitatively. Figure 1b illustrates the paradigm used to assay apparent motion during fixed flight. The results in Fig. 2 reveal the existence of a motion illusion in *Drosophila* which is strikingly analogous to the corresponding illusion in man. The illusions coincide with respect to the phase and direction of the oscillatory component at low modulation frequencies, and with respect to the direction of the continuous component at high modulation frequencies of the illumination. Maximum intensity and spectral composition of the light source do not seem to be critical in the range 0.003–300 cd m⁻².

In the fly the illusion originates within the network of bidirectional motion detectors, each comprising motion-specific interactions between neighbouring input units of the visual system⁸. The steady-state responses of the detectors are adequately described by the correlation model in Fig. 3, where the interaction is achieved by multiplication of two input signals. The model is representative of all 'simple' motion detectors which do not require more than the theoretical minimum of two input units in second-order nonlinear interaction⁹. Models of this type are equivalent in their steady-state responses, and seem to be sufficient to describe the properties of entirely different motion detection systems, such as the direction-selective retinal ganglion cells in vertebrates^{10,11}. The models respond to successive stimulation of their input units and cannot discriminate true motion from the corresponding ϕ -motion. Figure 3 shows the simultaneous sinusoidal stimulation of the receptors in the present paradigms. The visual input is characterised by the transmission-dependent mean intensities $\bar{I}_1 > \bar{I}_2$, and by the transmission-independent relative amplitude, or modulation, $\Delta I/\bar{I} = \Delta I_1/\bar{I}_1 = \Delta I_2/\bar{I}_2$. In the linear range of signal processing at the receptor level (N) the gain factor of the response, R , of the model is $\bar{I}_1\bar{I}_2(\Delta I_2/\bar{I}_2 - \Delta I_1/\bar{I}_1) = 0$. This is easy to show by application of the operations specified in Fig. 3. The absence of responses at the level of motion detection excludes the generation of the expected responses at subsequent levels. Nonlinear processing of the simultaneous signals before motion detection becomes necessary to elicit non-zero responses from the detector model.

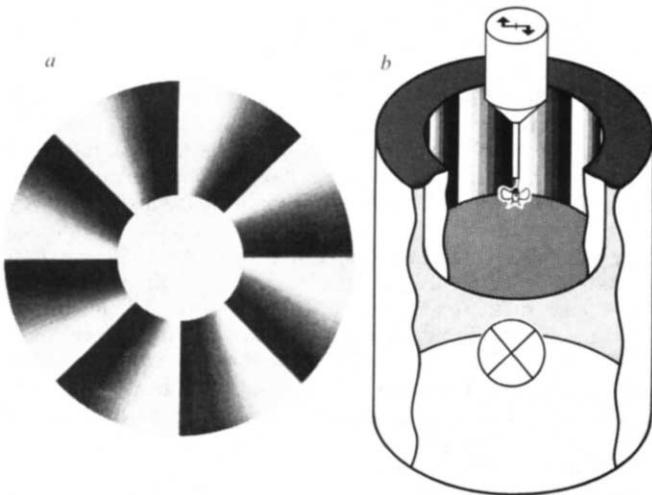


Fig. 1 *a*, Circular array of wedge filters used to elicit, stroboscopically, the illusion of oscillatory (0.3–2 Hz) or continuous (5–12 Hz) rotation in the human observer. The best results were obtained with a transparency of the pattern (absorbance 0–2), illuminated from behind by a sinusoidally modulated light source of almost arbitrary colour and intensity. It may be difficult to see the illusion when the figure is directly exposed to flickering light. *b*, Cylindrical array of wedge filters surrounding the fruitfly *Drosophila* during fixed flight on a torque meter which is used to record the course control response. The stationary pattern was illuminated from outside by a programme-controlled central light source in the lower compartment. The panorama was inverted between experiments and was presented in arbitrary angular positions to minimise lateral disparities of stimulus and response.

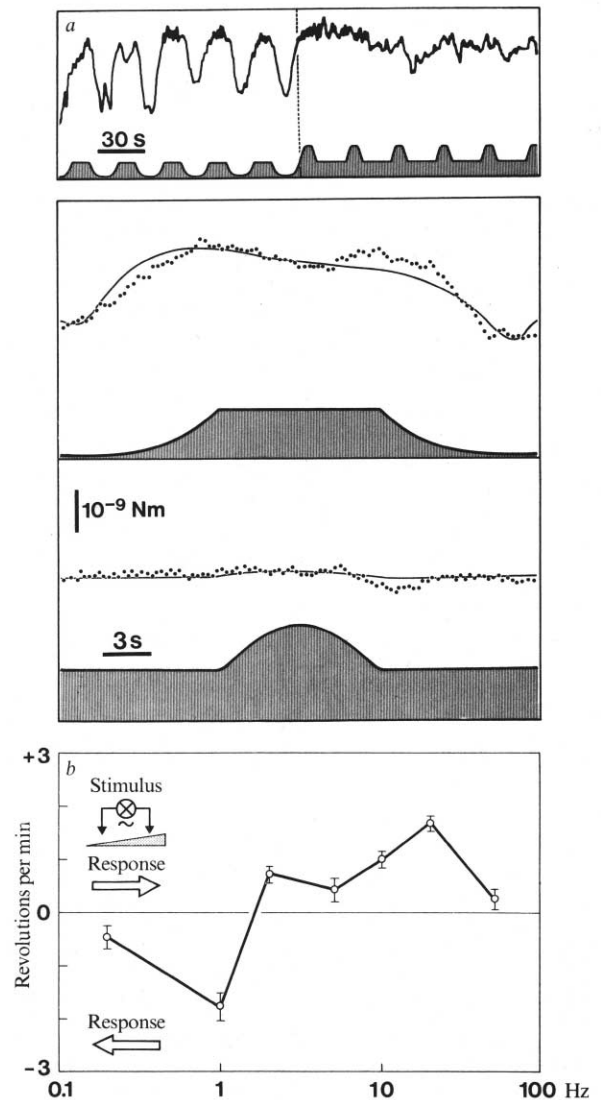


Fig. 2 *a*, Oscillatory motion illusion response of the fruitfly during fixed flight in the experimental set-up of Fig. 1b. The shaded area shows the actual light programme ($I_{\max} = 0.003$ cd m⁻²). The upper diagram shows the torque about the vertical axis produced in a single experiment. Increasing torque indicates the increasing tendency to turn towards the less transparent side of the filters, and vice versa. The dots in the lower diagrams refer to the responses of 12 flies, or the averages of 257 subsequent cycles of stimulation. The solid curves represent the motion illusion response of the model in Fig. 3. The oscillatory motion illusion of the fly is predictable and resembles the illusion induced in the human observer. *b*, Continuous motion illusion response of the stationary walking fruitfly as a function of the frequency of sinusoidal light modulation in the panorama of Fig. 1b. The tendency to turn was measured in revolutions per metre pathlength. The means and s.e.m. refer to the time averages of the responses of 11 flies which covered a total pathlength of 2,200 m in 250 h.

Surprisingly, the characteristics describing the apparent saturation of photoreceptors in quasi-static conditions, $X(t) \sim I(t)/(I(t) + I_0)$; $I_0 = \text{constant}$ (refs 12, 13), are apparently sufficient to elicit the expected oscillatory responses. To calculate R , the receptor output, $X(t)$, may be developed into a power series about $\bar{I}$. The first-order term accounts for a positive gain factor $I_0(\bar{I}_1 - \bar{I}_2)\Delta I/\bar{I} > 0$ and describes the illusion qualitatively. The results of a quantitative simulation are shown in the lower sections of Fig. 2a. The solid curves approximate the oscillatory responses of the fly with respect to (1) the light-on effect, (2) the light-off effect, (3) the phase lead on the periodic stimulus¹⁴, and (4) the decrease of the response with

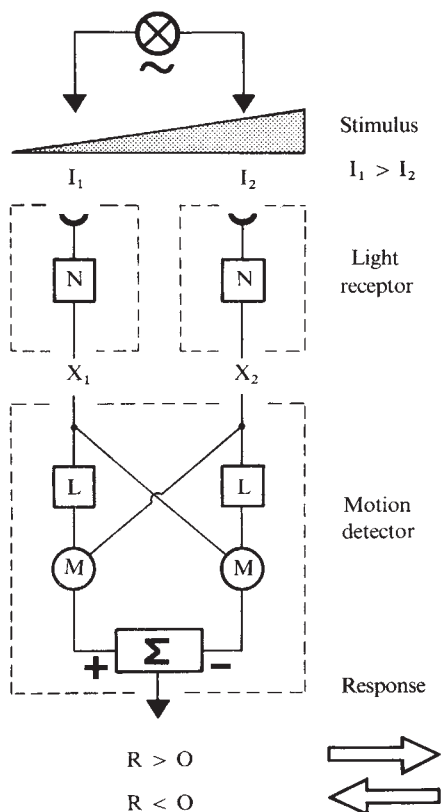


Fig. 3 Motion detector model used to describe the perception of both true and apparent motion in man and fly. The figure shows the attenuation of a sinusoidally modulated light source by a wedge filter. The simultaneous signals, I_1 and I_2 , are nonlinearly transformed (N) into the input signals, X_1 and X_2 , of a motion detector of the Reichardt-Hassenstein type⁹. The model consists of linear elements (L) for the retardation, and nonlinear elements (M) for the multiplication, of the input signals. It produces responses R of opposite sign if the direction of motion is reversed. Partition of the bi-directional motion detector into uni-directional antagonists imposes no constraints on the present results. The motion illusion obtained with the model is due to the nonlinear properties of N . The responses disappear if N becomes linear.

increasing background illumination. The matched parameters $I_0 = 0.04 I_{\max}$ of the receptor characteristics (N), and the matched time constant $T = 10$ s of the low pass filters (L), are also appropriate for the simulation of the response to true motion.

The simulated illusion decreases with progressing linearisation of the receptor characteristics at higher frequencies of stimulation. Concurrently, dynamic receptor nonlinearities become important, in particular, the prolonged latency of the light-on signal of the receptor output at lower levels of illumination^{15,16}. Latency differences between the light-on signals in the present experiments lead to successive stimulation of the input units of the motion detectors in the expected direction, towards the less transparent side of the wedge filters. The oscillatory motion illusion cannot be explained by prolonged latency. Qualitatively, there is no resemblance between the rebound of the transient light-on response $R > 0$ of the model and the marked light-off effect shown in Fig. 2a. Quantitatively, latency differences of 0–25 ms in man⁶, or 0–5 ms in the fly¹⁷, seem to be too small to contribute significantly to the oscillatory motion illusion³. However, accumulation and fusion of latency-induced transients $R > 0$ seem to account for the continuous motion illusion at higher frequencies of stimulation in Fig. 2b. The negative responses at medium frequencies are possibly due to the inversion of the brightness gradient at the edges between neighbouring wedge filters.

The results obtained permit the classification of various types of motion illusion according to their origin within the visual system. Unlike ϕ -motion² the present illusion requires nonlinear signal processing before motion detection, and is adequately, though not necessarily⁴, explained by the intensity dependence of gain and latency at the receptor level.

We thank E. Buchner, D. M. MacKay, T. Poggio, G. Westheimer and others for valuable comments.

HEINRICH BÜLTHOFF
KARL G. GÖTZ

Max-Planck-Institut für biologische Kybernetik,
D-7400 Tübingen, FRG

Received 16 October 1978; accepted 26 February 1979.

1. Wertheimer, M. *Z. Psychol.* **61**, 161–265 (1912).
2. Kolers, P. A. *Aspects of Motion Perception* (Pergamon, Oxford, 1972).
3. Alpern, M. A. *M.A. Archs Ophthal.* **51**, 369–374 (1954).
4. Andrews, D. P., MacKay, D. M. & Wilson, J. P. *Nature* **199**, 161 (1963).
5. Anstis, S. M. *Science* **155**, 710–712 (1967).
6. Lit, A. *Am. J. Psychol.* **62**, 159–181 (1949).
7. Götz, K. G. *Kybernetik* **2**, 77–92 (1964); **4**, 199–208 (1968).
8. Buchner, E. *Biol. Cybernetics* **24**, 85–101 (1976).
9. Reichardt, W. & Poggio, T. Q. *Rev. Biophys.* **9**, 311–438 (1976).
10. Barlow, H. B. & Levick, W. R. *J. Physiol., Lond.* **178**, 477–504 (1965).
11. Torre, V. & Poggio, T. *Proc. R. Soc.* **202**, 409–416 (1978).
12. Naka, K. I. & Rushton, W. A. H. *J. Physiol., Lond.* **185**, 536–555 (1966).
13. Dawis, M. *Biol. Cybernetics* **30**, 187–193 (1978).
14. Thorson, J. *et al. Science* **145**, 69–71 (1964); **183**, 161–172 (1974).
15. Baylor, D. A., Hodgkin, A. L. & Lamb, T. D. *J. Physiol., Lond.* **242**, 658–791 (1974).
16. Kramer, L. *et al. Biophys. Struct. Mechanism* **1**, 239–257 (1975); **2**, 333–336 (1977).
17. Gemperlein, R. & Smola, U. *J. comp. Physiol.* **78**, 30–52 (1972); **79**, 363–392 (1972).

Activity of an identified wind interneurone in a flying locust

LOCUST flight has been intensively studied by behavioural physiologists. The basic flight rhythm seems to be generated within the thoracic ganglia¹, although its timing is modified from wing beat to wing beat by phasic input from the wing sense organs². In this way, the basic motor score is possibly tuned to operate at the resonant, most energetically efficient frequency of the individual insect, and adjusted to accommodate changes in wing movements caused by variations of the external air current. Sensory influences on the flight motor neurones do not only originate at the wings. Weis-Fogh has demonstrated that stimulation of the wind-sensitive head hairs causes tonic excitation of these motor neurones³. Here, we have recorded in the locust during flight, the activity of two of the interneurons that convey wind information from the head hair afferents to the thoracic ganglia. These interneurons, the tritocerebral commissural giants (TCG)^{4,5}, receive their afferent wind hair input in the brain and make excitatory connections with flight motor neurones in the thoracic ganglia⁶. We show that these cells are rhythmically active during flight, synchronous with the wing beat, as a result of periodic stimulation of the head hairs. It is therefore possible that the wind-hair sensory system has a phasic influence on the flight motor as well as the already clearly established tonic influence.

We were able to record the activity of the TCG neurones during tethered flight because they take a curious path through the tritocerebral commissure, a nerve of small diameter (60 μ m) which loops under the gut and links the tritocerebral hemispheres of the brain (Fig. 1). With an axon diameter of 20 μ m, these interneurons are much larger than any of their neighbours in the commissure. This situation was exploited to make long-term recordings of the TCG at the commissure, using a chronically implanted hook electrode.

We used adult male *Locusta migratoria* from laboratory culture, selecting animals that demonstrated a strong willingness to fly. Insulated steel wires (diameter 30 μ m) were used to record from the interneurone and hind wing depressor muscle 129 in the thorax. Animals were supported at the sternum near

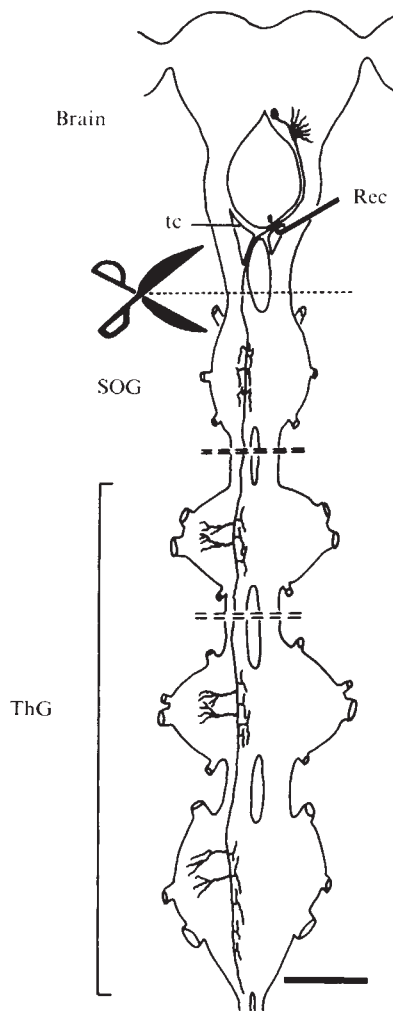


Fig. 1 Schematic view of the CNS in the head and thorax of the locust to show the structure of one of the paired TCG wind interneurons. The TCG is seen to pass through the tritocerebral commissure (tc) at which point it can be recorded with a hook electrode (Rec). The dotted line indicates the position at which the connectives were cut (see text). SOG, suboesophageal ganglion; ThG, thoracic ganglia. Scale bar, 0.5 mm.

the mouth of a wind tunnel producing a laminar air stream of 3.5 m s^{-1} .

Figure 2a shows the low level of spikes in the TCG interneurone in a resting animal exposed to a laminar air stream. In contrast, during flight the neurone displayed a much higher level of spiking activity which was rhythmical (Fig. 2b) and synchronised with the firing of a wing depressor muscle in the thorax (Fig. 2c). This effect persisted in animals flying in the dark, thus excluding the possibility that it arises from visual input produced by the movements of the wings.

The rhythmicity in the TCG neurone during flight may occur simply because the insect can sense phasic changes in turbulence around the head created by its own wing or head movements. Alternatively, the cell may receive rhythmical input from the flight motor or sense organs in the thorax, neck or abdomen. As the TCG has been observed to carry only descending information, such input would first have to reach the brain through ascending interneurons.

To investigate these two possibilities, we recorded the activity of the cell in an animal in which we had cut both circumoesophageal connectives behind the tritocerebral commissure (dotted line, Fig. 1). In these conditions of complete cessation of excitatory input from the head hairs to the flight motor neurones

in the thorax, flight was difficult to initiate and the wing beats were irregular. However, in a few cases continuous flight lasting up to 5 min could be induced. All lesions were verified after the experiment.

Figure 3 shows the results for an animal flying straight ahead. The activity of the cell was much diminished as a result of the lesion (Fig. 3a) but the infrequent spikes that occurred were synchronised with the firing of a wing depressor muscle in the thorax (Fig. 3b, c). Removing the antennae had no effect on this synchronisation.

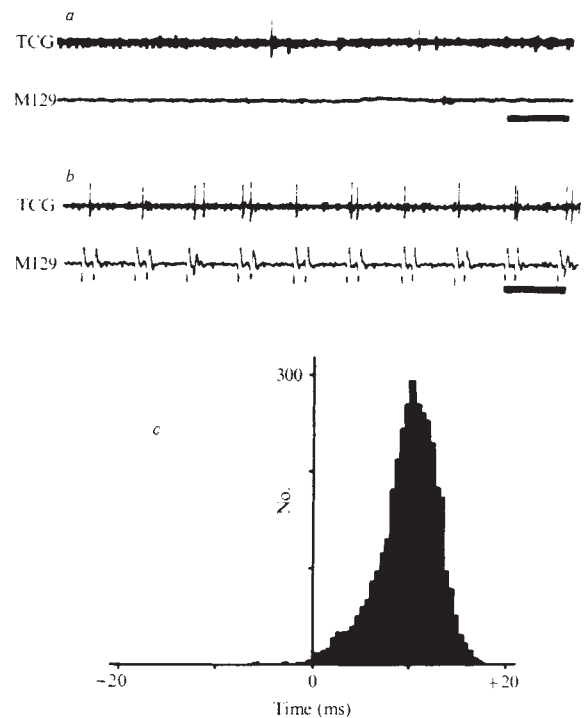


Fig. 2 Activity of the TCG interneurone (top trace) and wing depressor muscle 129 right (bottom trace) in a locust suspended in a laminar air stream. *a*, The weak response of the TCG in a resting animal. *b*, During straight flight the neurone shows rhythmical activity. Timescale, 50 ms. *c*, Histogram showing time of firing of the TCG in relation to the activity of the muscle. Abscissa: time of occurrence of interneurone spike relative to muscle spike; ordinate: number of interneurone spikes. The histogram is a record of 3,695 TCG spikes sampled during 3,500 successive wing beats. Firing of the TCG and muscle are well synchronised.

This experiment demonstrated clearly that the TCG receives rhythmical input from the wind-sensitive hairs on the head. In addition, the lower level of firing of the TCG in this preparation, compared with that in the intact animal, showed that ascending information is also a significant input to the cell. However, other experiments in which interneurone recordings were made from animals with their heads completely waxed over (but connectives intact) showed that this ascending input contributes none of the rhythmicity during flight. We therefore conclude that the flying locust is sensitive to the phasic air currents caused by the rhythmical flight movements of its own body.

Intracellular recording has revealed that this interneurone makes functionally direct excitatory connections with flight motor neurones in the thoracic ganglia⁶. Therefore these motor neurones receive a phasic excitatory input, through the TCG, from the wind-sensitive hairs during flight. It remains to be seen, however, whether the phasic information transmitted by the TCG (and presumably by other wind-sensitive interneurons) is a behaviourally significant fast-feedback loop for the flight

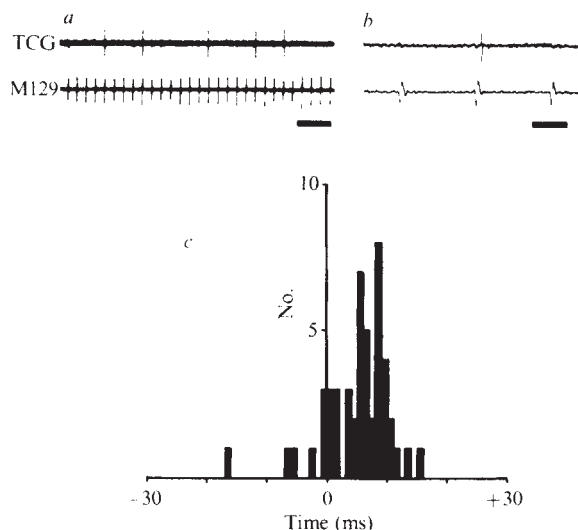


Fig. 3 Results obtained during straight flight in an animal in which both circumoesophageal connectives had been cut posterior to the tritocerebral commissure. *a*, Recordings of the interneurone (top trace) and the wing depressor muscle 129 right (bottom trace). Timescale, 200 ms. *b*, Expanded time axis showing interneurone spike in time with the contraction of the muscle. Timescale, 20 ms. *c*, Histogram showing time of firing of the interneurone in relation to the activity of the muscle, axes as for Fig. 2c. The data are a record of 48 TCG spikes sampled during 772 successive wing beats. Firing of the cell is fairly well synchronised with the contraction of the wing depressor muscle.

motor. Certainly, descending light-sensitive interneurons, when stimulated with a stroboscopic light, can entrain the flight frequency⁷. Perhaps some visual interneurons, able to detect the up-and-down movements of the wings, and wind-sensitive interneurons, work with the sensory inputs from the wings to set the optimum flight frequency for the insect. Alternatively, the phasic excitation that arises at the wind hairs, when conducted down the cord by many interneurons working in parallel, may arrive at the flight motor neurones as an integrated tonic excitation, and so function simply as a positive feedback mechanism to maintain flight. Further investigations using this preparation present exciting possibilities for a more detailed analysis of the neural basis of flight.

We thank Professors F. Huber and W. Nachtigall and Dr J. Altman for reading the manuscript. Drs W. Zarnack and H. K. Pfau discussed the results with us. The work was done using apparatus provided by a grant from the Stiftung Volkswagenwerk to Professor Nachtigall. J.B. is in receipt of a Max Planck Stipendium.

JON BACON

Max-Planck-Institut für Verhaltensphysiologie,
Abt. Huber,
D-8131 Seewiesen, FRG

BERNHARD MÖHL

Arbeitsgruppe Nachtigall im Fachbereich Biologie
der Universität des Saarlandes,
D-6600 Saarbrücken, FRG

Received 5 June 1978; accepted 20 February 1979.

1. Wilson, D. M. *J. exp. Biol.* **38**, 471–490 (1961).
2. Wendler, G. *J. comp. Physiol.* **88**, 173–200 (1974).
3. Weis-Fogh, T. *Nature* **164**, 873–874 (1949).
4. Bacon, J. thesis, Univ. Manchester (1977).
5. Bacon, J. & Tyrer, M. *J. comp. Physiol.* **126**, 317–325 (1978).
6. Bacon, J. & Tyrer, M. *Naturwissenschaften* **66**, 116–117 (1979).
7. Waldron, I. *Z. vergl. Physiol.* **57**, 331–347 (1968).

The segmentation of axons from the segmental nerve roots to the chick wing

THE chick wing is innervated by adjacent spinal roots which meet at the brachial plexus and then redistribute their axons to the peripheral nerves. Both the anatomical arrangement of nerves within the wing and the innervation territories supplied by each root are constant among animals of the same strain^{1–3}. We describe here the pathways taken by the axons from each of the main roots to their particular territories in the normal wing, and also in wings which have dorso-ventral and antero-posterior axes reversed relative to the host axes. Our observations suggest that the characteristic innervation territories of the segmental roots may arise from two basic rules: that axons travel in parallel over long distances without crossing, and that the branching pattern of nerves within the limb is determined by the local environment^{3,4}.

The motor and sensory innervation territories of the three segmental roots were mapped electrophysiologically *in ovo* at 17 d of incubation. Sensory innervation maps were constructed by mechanically or electrically stimulating localised regions of skin while recording from each nerve root in turn (Fig. 1). The motor innervation was mapped by stimulating each root in turn

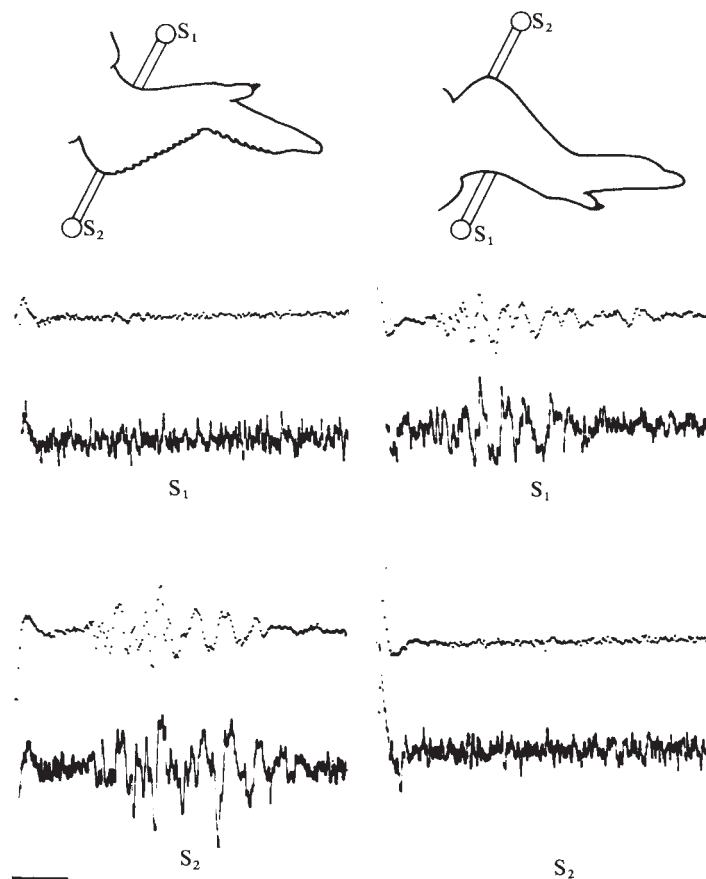


Fig. 1 Responses, recorded from the peripheral cut end of n16 using suction electrodes, to electrical stimulation of the skin in the normal (to the left) and rotated (to the right) wings of 578. Positions of the bipolar stimulating electrode (0.5-ms pulse) are shown at the top of each column. The top trace in each pair shows the averaged response to 16 consecutive stimuli, the bottom trace the response to a single pulse. Calibration bar, 10 ms.

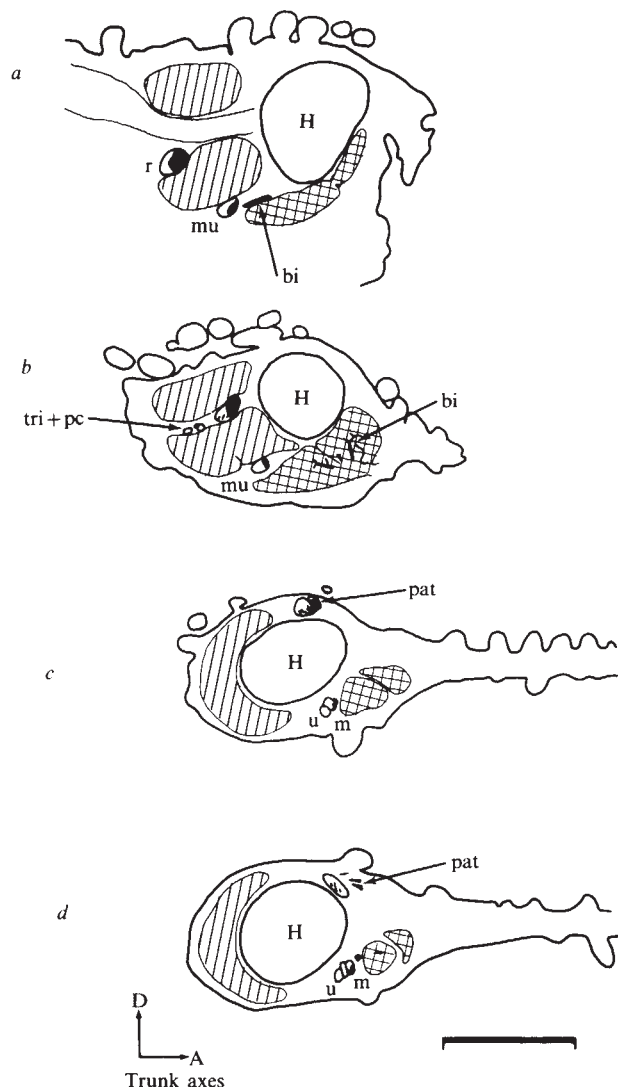


Fig. 2 Normal wing, n14 filled. Camera lucida drawings of transverse sections at four levels, all viewed from the elbow. Filled axons are shown in black; they occupy the anterior sectors of the radial and medio-ulnar nerves. Trunk axes shown; A, anterior; D, dorsal; calibration bar, 1 mm. *a*, Biceps nerve contains axons from n14 (arrowed). *b*, Triceps nerve (arrowed) has no filled axons, but n14 axons are clearly visible in the biceps muscle (arrowed). *c*, Filled axons leave the radial nerve at the patagiales nerve. A photograph of this is shown in Fig. 4a. *d*, Patagiales nerve with filled axons going to the web of the wing (arrowed). Biceps muscle cross-hatched, triceps muscle hatched. LAT DORS, latissimus dorsi muscle; EMR, extensor metacarpalis radialis; H, humerus; r, radial nerve; tri, triceps nerve; pc, posterior cutaneous nerve; m, median nerve; u, ulnar nerve; pat, patagiales nerve.

while observing electrical events accompanying muscle contraction. In all animals of our strain (White Leghorns), the 14th segmental root (n14) innervates the proximal anterior region including the biceps muscle, and the 16th segmental root (n16) supplies the posterior and distal regions including triceps. The territory of n15 lies in between and overlaps both³.

In our experimental animals the wing buds were rotated 180° at stage 19, a significant time before the axons invade the bud. These limbs develop autonomously^{5,6} with dorso-ventral and antero-posterior axes reversed with respect to the host. For simplicity we will refer to the axes of trunk and wing separately; thus, in control animals these axes are the same and in operated animals the wing axes are opposite to those of the host trunk. When we map the sensory and motor innervation pattern of these animals we find that n14 now goes to the posterior and

distal parts of the wing (including triceps and the hand), whereas n16 innervates the anterior region of the grafted wing (including biceps and the alar web). The pattern of innervation of the rotated limbs is therefore (within the limits of variation found in control animals) the converse of that in the normal wing (Fig. 1).

Dissection of animals after recording shows that the spinal roots meet to form a seemingly normal plexus in the trunk which

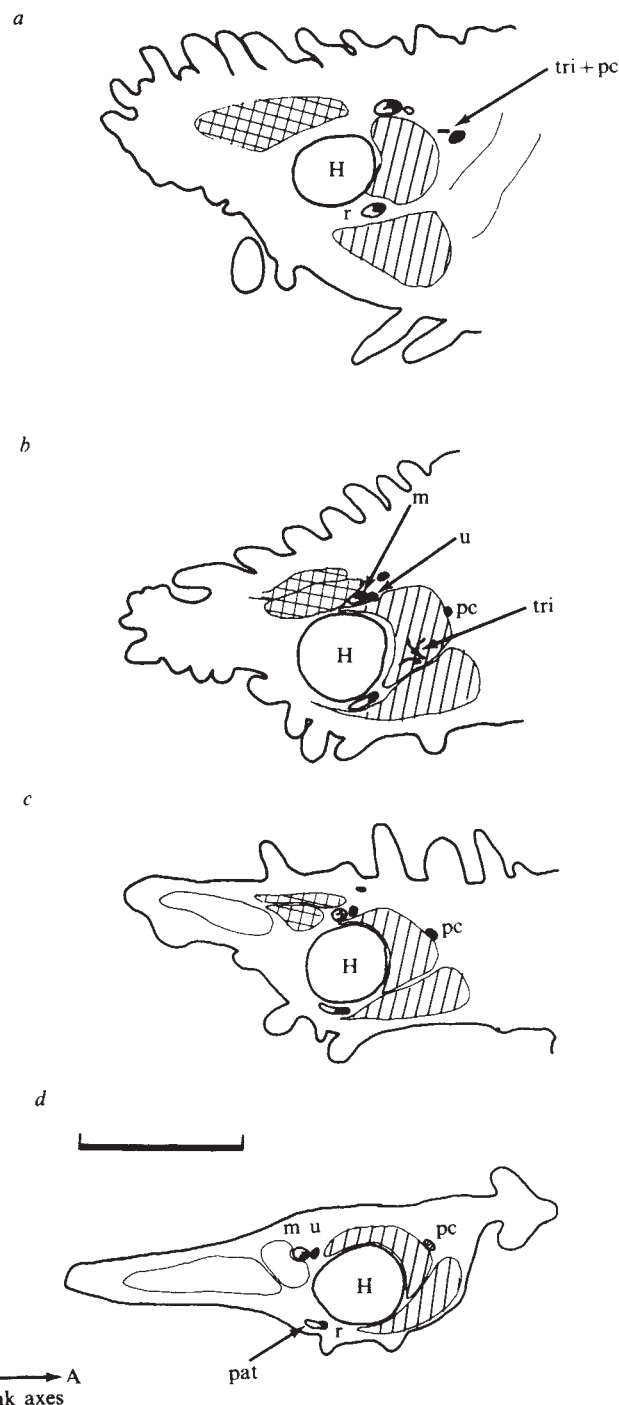


Fig. 3 Rotated wing, n14 filled. Camera lucida drawings of transverse section at four levels, all viewed from the elbow. Trunk axes are as shown, wing axes are the reverse, that is, anterior to the left and dorsal downwards. Filled axons are shown in black; they occupy the posterior sector of nerves in the graft. *a*, Triceps and cutaneous nerves with filled axons (arrowed). *b*, Triceps muscle contains filled axons (arrowed), biceps does not. Filled axons are also present in median and ulnar nerves (arrowed). *c*, Fascicle of patagiales nerve begins to separate from radial nerve. *d*, Unfilled axons leave the radial nerve to form patagiales nerve (arrowed). A photograph of this is shown in Fig. 4c. Abbreviations as for Fig. 2.

produces the normal main dorsal and ventral branches to the wing. On entering the rotated graft the dorsal branch apparently behaves as if it were the ventral branch forming the medio-ulnar nerve. The ventral branch behaves as if it were the dorsal branch forming the radial nerve. Thus, the anatomical arrangement of nerves within the graft is very similar to that found in the normal wing.

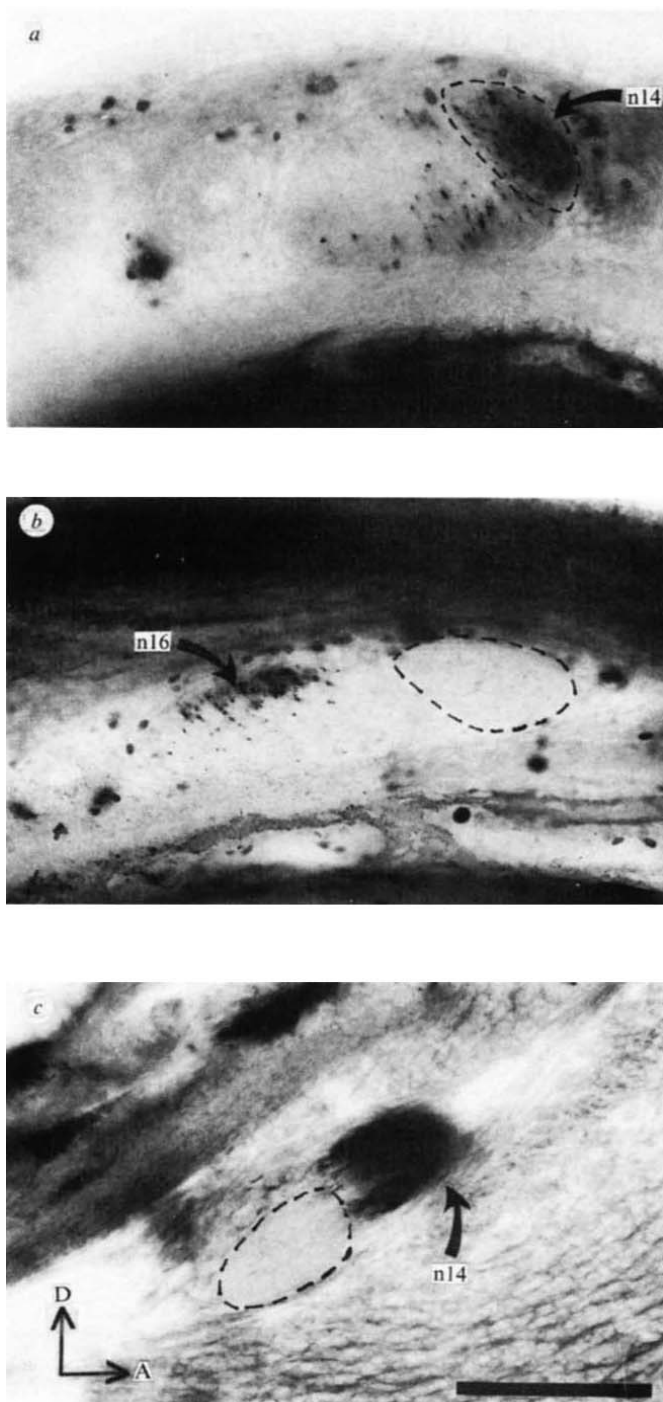


Fig. 4 Photomicrographs of transverse sections of radial nerve including the branch point of the patagiales nerve. The arrows at the bottom indicate trunk (host) axes for all photographs; D, dorsal; A, anterior. The sector surrounded by the dotted line forms the patagiales nerve. The arrows indicate filled axons. Calibration bar, 100 μ m. *a*, Radial nerve from section shown in Fig. 2c, n14 filled. *b*, Radial nerve from section taken from a similar level to Fig. 4a, in which n16 had been filled. *c*, Radial nerve from section shown in Fig. 3d. There are no filled fibres in the fascicle giving rise to the patagiales nerve although n14 was filled.

The anatomy and physiology of the rotated wings suggest that although the pattern of nerves is determined by the environment, the arrangement of axons within the nerves is non-selective. To test this directly we have traced the connections anatomically by filling the axons of each root in turn with cobalt chloride, which can subsequently be visualised after treatment with H_2S and silver intensification⁷⁻⁹. When we fill n14 in the normal wing, we find that its axons maintain their anterior position through the plexus and into the main dorsal and ventral nerves, whereas axons from n16 maintain a posterior position, and axons from n15 lie in the region between the two and overlap both. The distribution of axons within the nerves as they enter the wing is therefore related to their spinal root origin. We further find that the axons continue to run in parallel, maintaining their positions over long distances, certainly to below the elbow, so that the cross-sections of each nerve consist of sectors which have relatively homologous axonal origins (Fig. 2). When a branch leaves the main nerve from its anterior aspect it contains axons which have been running in the anterior sector (Fig. 2c, 4a). When a branch leaves the main nerve posteriorly it contains axons from the posterior sector (Fig. 4b). Thus, n14 axons enter branches to the alar web and the biceps, whereas n16 enters posterior branches such as the triceps and ulnar nerves.

In the rotated wings the axons of n16 enter the brachial plexus in the trunk in their normal posterior position with those of n14 in their normal anterior position. This arrangement is maintained through the plexus and into the graft (Fig. 3). Thus, the axons of n14 come to occupy posterior sectors of the wing dorsal and ventral nerves, and are then channelled into wing posterior branches, for example, triceps and ulnar nerves (Figs 3, 4c). Conversely, axons from n16 from their posterior sectors in the trunk nerves come to occupy anterior sectors of the wing nerves and travel down anterior branches to the alar web and biceps.

There is no accepted theory to explain how each segmental root comes to innervate its particular limb territory. Recently, ideas have polarised to two extremes: target specificity where axons actively seek out their 'appropriate' targets, or initial random outgrowth with 'inappropriate' axons being subsequently eliminated by cell death¹⁰⁻¹⁴. Our results suggest that the pattern of innervation results from the way axons travel together in parallel over relatively long distances and the manner in which the segmental roots combine together in the plexus. The fate of individual axons (once they have taken up a certain position in the plexus) is determined by the peripheral branching pattern, and this is clearly controlled by the environment rather than by the nervous system. We do not have any information on the spinal cord anatomy and the next stage of our investigation is to observe the motor neurone pools supplying such rotated grafts.

R. VICTORIA STIRLING
DENNIS SUMMERBELL

National Institute for Medical Research,
The Ridgeway,
Mill Hill,
London NW7, UK

Received 29 November 1978; accepted 9 March 1979.

1. Roncali, L. *Monitore zool. ital.* (N5) **4**, 81 (1970).
2. Landmesser, L. & Morris, D. *J. Physiol., Lond.* **249**, 301 (1975).
3. Stirling, R. V. & Summerbell, D. *J. Embryol. exp. Morph.* **41**, 189 (1977).
4. Hamburger, V. *Physiol. Zool.* **12**, 268 (1939).
5. Summerbell, D., Lewis, J. H. & Wolpert, L. *Nature* **244**, 492 (1973).
6. Summerbell, D. & Lewis, J. H. *J. Embryol. exp. Morph.* **33**, 621 (1975).
7. Pitman, R. M., Tweedle, C. D. & Cohen, M. J. *Science* **176**, 412 (1972).
8. Bacon, J. P. & Altman, J. S. *Brain Res.* **138**, 359 (1977).
9. Stirling, R. V. *J. Physiol., Lond.* **280**, 3-4P (1978).
10. Hughes, A. & Prestige, M. C. *J. zool. Res.* **152**, 347 (1967).
11. Prestige, M. C. *J. Embryol. exp. Morph.* **18**, 359 (1967).
12. Landmesser, L. & Morris, D. *J. Physiol., Lond.* **249**, 301 (1975).
13. Lamb, A. H. *Dev. Biol.* **54**, 82 (1976).
14. Pettigrew, A. G., Lindemann, R. & Bennett, M. R. *J. Embryol. exp. Morph.* **49**, 115 (1979).

Single glutamate-activated channels in locust muscle

EVIDENCE indicates that glutamate-activated membrane channels (glutamate channels) are responsible for the post-synaptic potential at insect¹ and crustacean² excitatory neuromuscular junctions. They may also be involved in the regulation of pre- and postsynaptic activity in the central nervous system of vertebrates^{3,4}. Analysis of the current noise produced by application of glutamate to excitatory synapses on locust muscle suggested that glutamate channels have a conductance of about 125 pS and a mean open time of about 2 ms (refs 5, 6). Direct measurements of the currents passing through individual acetylcholine-activated channels (ACh channels) in

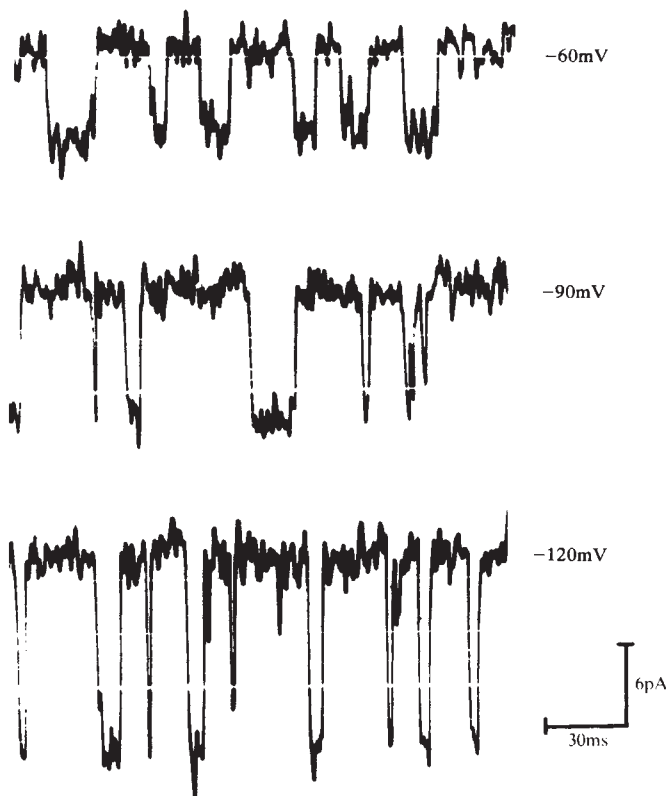


Fig. 1 The currents through single glutamate-activated channels in the metathoracic extensor tibiae muscle of a 6-d denervated locust (*Locusta migratoria*). The muscle cell was voltage clamped with standard techniques, and the membrane potential held at the values shown next to each trace. A patch electrode filled with physiological saline containing 100 μ M L-glutamate was pressed on to the surface of the cell, and measured the currents passing through the muscle membrane under the patch. The electrode had a 2- μ m fire-polished tip which had a free-solution resistance of about 2 M Ω . When pressed on to the cell, the resistance increased to more than 10 M Ω , electrically isolating the membrane patch with a voltage divider ratio (leak to series resistance) of more than 5. The downward events are currents passing from the electrode into the cell. The background noise is the Johnson noise expected for the combination of the series and leak resistances. The patch electrode was placed within a cell diameter of the intracellular voltage electrode. Currents measured from the patch electrode were amplified and stored on an analogue FM tape recorder. The illustrations were made by play-back of the data through an active 4-stage filter ($f_c = 450$ Hz) with Bessel characteristics. All three records were from the same patch. The temperature was 21 °C. Standard locust physiological saline was used for all experiments. The cell had been previously treated for 15 min with 1 μ M Con A in saline, then washed with normal saline for more than 30 min. No single channel responses could be measured when the glutamate was omitted from the patch electrode.

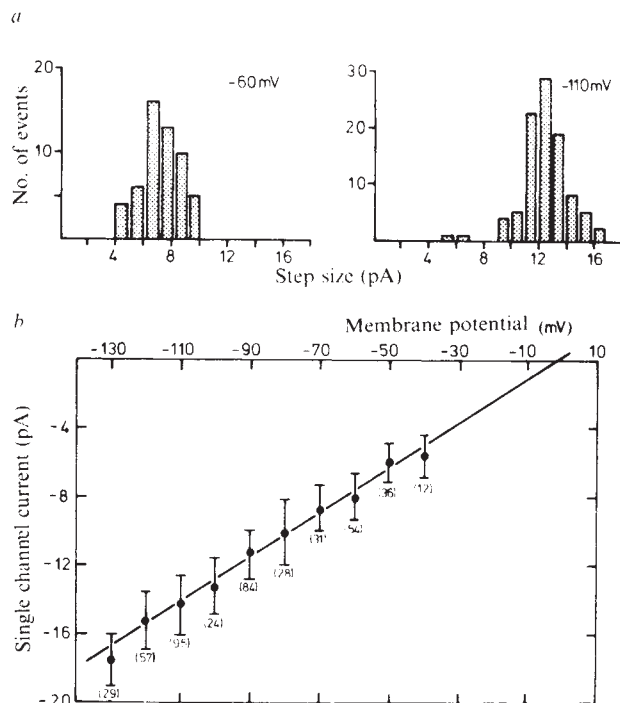
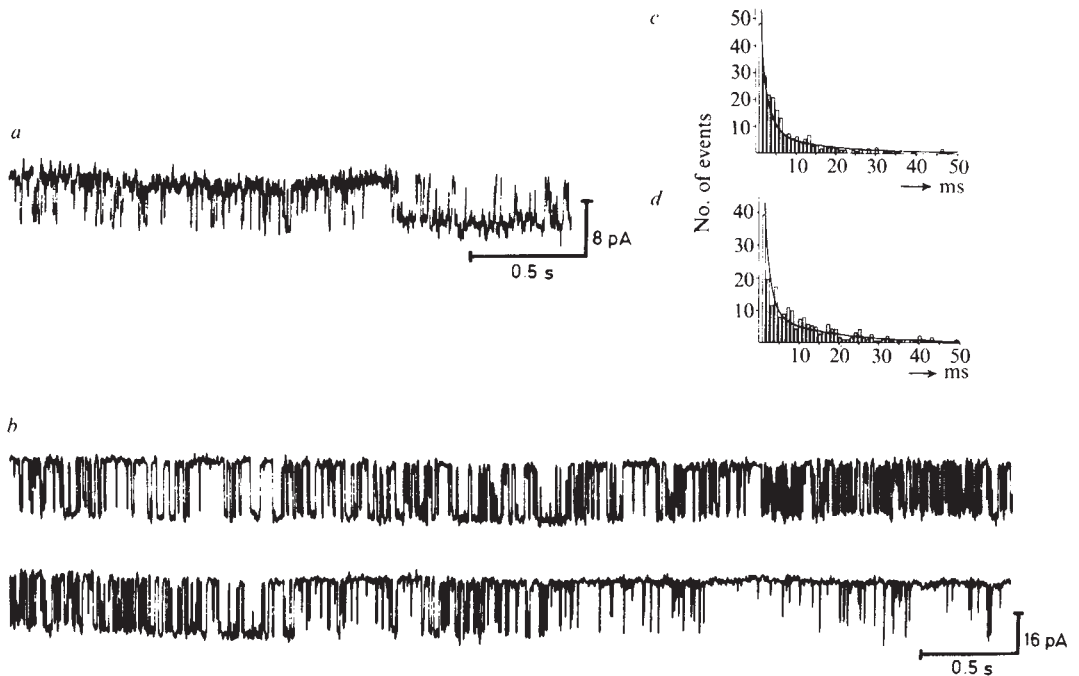


Fig. 2 Current-voltage characteristics of single glutamate channels. Patch recordings at various different membrane potentials were sampled with a PDP 11/34 computer, and channel amplitudes measured as the difference between the average value of the current 2 ms before and after individual current steps. *a*, Histograms of measured step sizes at two membrane potentials, -60 and -110 mV. Channel amplitudes measured by hand from strip chart recordings gave essentially the same distributions as those illustrated here. *b*, The mean of the measured step sizes and the standard deviations plotted as a function of the holding potential. Numbers in brackets are the number of individual events analysed at each potential. The solid line is the least-squares fit to the mean value at each holding potential. The slope of this line was 133 pS. The zero current intercept was -2.5 mV. The recording conditions were the same as described in Fig. 1.

vertebrate muscle membrane have recently been made. The single channel conductance (about 25 pS) and kinetics correspond well with those inferred from noise analysis^{7,8}. We report here the measurement of the currents passing through individual glutamate channels in locust muscle. The conductance of an open single channel is about 130 pS. The gating kinetics are complex, and evidence is given for sudden changes in an individual channel's distribution of open and closed times.

Single glutamate channels have been measured on metathoracic extensor tibiae and retractor unguis muscles in adult locusts (*Locusta migratoria*) by placing a patch electrode containing low concentrations of glutamate against the surface of the cell. However, due to a low channel population density and rapid desensitisation of extrajunctional D receptors, these records were difficult to obtain and often contained less than 10 individual events. The population density of extrajunctional D receptors was increased by using muscles denervated for at least 6 d (refs 9–11). Nevertheless, even with denervated muscle, desensitisation was often complete before the patch electrode could be properly sealed against the membrane. Therefore, most studies were made on extrajunctional receptors of chronically denervated muscle which was pretreated with concanavalin A (Con A). Brief exposure of the cells to this lectin irreversibly prevents the desensitisation of junctional glutamate and extrajunctional glutamate D receptors on locust muscle^{12,13}. Pretreatment of the muscles with 1 μ M Con A in physiological saline for 15 min, followed by a thorough rinse, enabled us to make patch recordings which exhibited a stable level of channel activity for up to 30 min or longer.

Fig. 3 Complex kinetics of individual channels. *a*, A sudden kinetic state change in the activity record of what seems to be one channel. This record was made at the resting potential of this cell, -45 mV, without voltage clamp. The cell was from a 6-d denervated animal, and had been Con A treated. The mean channel open time before the transition was 2.8 ± 0.3 ms. After the transition it was 22.5 ± 4 ms. The probability that these events were from the same population, but distributed by chance, was $P < 0.001$. The mean closed times were 16.6 ± 2.1 ms, and 3.0 ± 0.5 ms, corresponding to a $P < 0.001$. *b*, a 6-s record of a channel's activity from a cell voltage clamped at -120 mV. The two segments



are contiguous. A broad variety of kinetic states are represented in this record. Note especially the long segment with a high frequency of fast activity, and the change to the ground state at the end of the segment. Both *a* and *b* were made by replaying pre-filtered data at a 60 times slower tape speed and recording on a normal strip chart recorder. f_c was 350 Hz. The reduced amplitude of some events was due to filter attenuation. Note the difference between the calibration scales in *a* and *b*. *c*, Histogram of the channel open times from the data segment in *b*. Open times were measured from digitised data as the time difference between the first downward deflection of the current and the next following upward deflection. All pattern recognition and signal-to-noise judgments were made by eye. Only transitions greater than 2–3 times the noise amplitude were measured. The data were unfiltered. The histogram was fitted with two exponentials using a non-linear, least-squares fitting program. The best fit was obtained with $\tau_{\text{fast}} = 1.8$ ms, $\tau_{\text{slow}} = 12.3$ ms and an amplitude ratio of 7.0. The normalised χ^2 for this fit was 1.0. *d*, Histogram of channel closed times made concurrent with *c*. The best fit was obtained with $\tau_{\text{fast}} = 1.6$ ms, $\tau_{\text{slow}} = 16.3$ ms, and an amplitude ratio of 6.2. χ^2 was 1.2.

Figure 1 shows glutamate channels in a 6-d denervated muscle which had been treated with Con A. The downward events are the sudden increase and subsequent decrease of inward currents. These currents are due to the opening and closing of individual channels. The current which flows through the open channel is essentially constant and depends on the membrane potential. At -60 mV, the approximate resting potential of the cell, this current is about 8 pA. At -120 mV, the single channel current is almost twice this value.

Single channel amplitudes are uniform in most cases. Uncertainties of measurement of these amplitudes are caused by the limited frequency response of the system and by the background noise, leading to some spread in the measured values. Figure 2*a* shows histograms of the channel currents at two different membrane potentials. The distribution centres roughly around a peak value, as in a gaussian distribution. Figure 2*b* is a plot of single channel current levels against membrane potential.

The conductance of the open channel seems to be independent of the membrane potential within the range of our measurements, as might be expected of a simple resistor. The points in Fig. 2*b* are fitted well by a straight line, whose slope corresponds to the conductance of the channel. The least-squares fit to the data points yields a line which has a slope of 133 pS and which has a zero-current potential very close to 0 mV. The extrapolated zero-current potential corresponds well with that of the miniature excitatory junctional currents (m.e.j.cs) in this cell. The m.e.j.cs have been shown to reverse at 0 to +5 mV, as do the responses to iontophoretically applied glutamate¹⁴.

The kinetics of the channel's open-close reaction are not stationary in these conditions. This inconstancy is most clearly observed during recordings where the frequency of the simultaneous opening of two or more channels (double events) is very low or zero. When two independent channels are open simultaneously within the measurement area the current measured should be double the normal current. The frequency of double

events is predicted by a binomial distribution, and for the high levels of activity illustrated here, this frequency is substantial. As few double events are observed, it is most probable that only one channel is active within the patch area. Recordings made in these conditions show that the channels can sometimes shift from one kinetic 'state' to another. Figure 3*a* shows one such transition. Both the mean open time and the mean closed time have suddenly shifted. Neither is likely to have come from the same exponentially distributed interval population found before the transition (see legend to Fig. 3). This segment is a small part of a recording lasting several minutes which contains virtually no double events. Transitions such as those in Fig. 3*a* occur very frequently in our records, and have been observed at several different membrane potentials.

Several different kinetic states have been observed in long records of what seems to be the activity of only one channel. The most common state is that shown in the beginning of the trace in Fig. 3*a*. In more than 90% of our measurements the activity was similar to that illustrated, that is, opening events with short (<5 ms) durations, separated by relatively longer (>20 ms) closed intervals. Histograms have been made of channel open times in long segments of data which seem to be only in this 'ground' state, and these open times are roughly exponentially distributed. However, the range of time constant values for the best fit exponentials vary from 1 to 3 ms, in many cases in identical conditions. These values are similar to those measured by noise analysis of the synaptic receptor kinetics and by the decay of m.e.j.cs⁵.

During records in which a channel's activity is not in the ground state, both the open and closed times are multi-modal. Figure 3*b* is an example of this type of activity. As was the case in the record shown in Fig. 3*a*, the number of double events was very low for several minutes before and after the illustrated record. Both the frequency of events and the event duration change many times during this segment, with an apparent return to the ground state at the end of the record. Figure 3*c* and *d* are

histograms of the open and closed intervals during this burst. Both histograms could only be adequately fitted by curves consisting of at least two exponentials. The lines shown are the least-squares fit to the data by double exponential curves. The various time constants are given in the legend to Fig. 3. The data indicate that the kinetics of this channel are the result of a complex process. Models of this channel's activity must take this into account. Furthermore, this multi-modal activity does not predict the relatively prolonged kinetic state changes observed in the channel's activity. These state changes must also be accounted for in a complete model of glutamate channel kinetics. It is not yet possible to determine whether the presence of Con A is responsible for the appearance of these multiple states. In several records made without Con A treatment, the channels seemed to be in the ground state. However, each of these records was too short, due to rapid desensitisation, to demonstrate the absence of other states without Con A.

In conclusion, direct measurement of the currents which pass through single glutamate-activated membrane channels in the locust muscle demonstrates that these channels have a conductance of about 130 pS and multiple kinetic states of channel gating. This conductance is almost five times larger than that of the ACh channels at the vertebrate neuromuscular junction, and is among the largest yet found in biological membranes. The measured conductance agrees well with that inferred from previous noise analysis in this system⁶. Our measurements also directly demonstrate that the conductance of the open channel is independent of voltage between -130 and -40 mV, and that the extrapolated reversal potential is very close to 0 mV, as has been suspected for this channel. We also measure a kinetic state of this channel which is very common, and which has a distribution of channel open times similar to those found for synaptic channels. Other kinetic states also occur in individual channels, often with sudden inter-state transitions. It has not been possible to determine if these states are also a property of the normal synaptic glutamate channel, or if they are induced either by the presence of Con A or by denervation. In either case, the observation that a single channel can suddenly change its kinetic properties is new, and must be investigated further. The patch electrode technique is the only method with resolution sufficient to clarify this problem further.

We thank Drs B. Sakmann and E. Neher for discussions and comments on the manuscript, and Dr R. Ramsey for technical support. This work was supported by a grant from the SRC to P.N.R.U. J.B.P. is supported by a post-doctoral fellowship from the Muscular Dystrophy Association of America.

J. B. PATLAK

Max Planck Institut für biophysikalische Chemie,
D-3400 Göttingen-Nikolausberg,
Am Faßberg, Postfach 968, FRG

K. A. F. GRATION
P. N. R. USHERWOOD

The University of Nottingham,
Department of Zoology,
Nottingham, UK

Received 13 December 1978; accepted 1 March 1979.

1. Usherwood, P. N. R. & Cull-Candy, S. G. in *Insect Muscle* (ed. Usherwood, P. N. R.) 207-280 (Academic, New York, 1975).
2. Crawford, A. C. & McBurney, R. N. *J. Physiol., Lond.* **258**, 205-226 (1976).
3. Davidson, N. & Southwick, C. A. P. *J. Physiol., Lond.* **219**, 689-708 (1971).
4. Curtis, D. R., Phillis, J. W. & Watkins, J. C. *Nature* **183**, 611-613 (1959).
5. Anderson, C. R., Cull-Candy, S. G. & Miledi, R. *Nature* **261**, 151-153 (1976).
6. Anderson, C. R., Cull-Candy, S. G. & Miledi, R. *J. Physiol., Lond.* **282**, 219-242 (1978).
7. Neher, E. & Sakmann, B. *Nature* **260**, 799-802 (1976).
8. Neher, E., Sakmann, B. & Steinbach, J. H. *Pflügers Arch. ges. Physiol.* **375**, 219-228 (1978).
9. Usherwood, P. N. R. *Nature* **223**, 411-413 (1969).
10. Cull-Candy, S. G. *Nature* **258**, 530-531 (1975).
11. Cull-Candy, S. G. *J. Physiol., Lond.* **276**, 165-181 (1978).
12. Mathers, D. & Usherwood, P. N. R. *Nature* **259**, 409-411 (1976).
13. Mathers, D. & Usherwood, P. N. R. *Comp. Biochem. Physiol.* **59C**, 151-155 (1978).
14. Anwyl, R. & Usherwood, P. N. R. *J. Physiol., Lond.* **242**, 86-87P (1974).

Effects of oxotremorine demonstrate presynaptic muscarinic and dopaminergic receptors on motor nerve terminals

THERE has been little conclusive evidence about the presynaptic effects of acetylcholine (ACh) and cholinomimetics on mammalian motor nerve terminals^{1,2} which may have functional implications in motor abnormalities^{3,4}. We now report the existence of dopaminergic muscarinic receptors on motor nerve terminals which participate in the local release of ACh from motor nerve terminals.

Details of the collection and assay of ACh released from isolated rat phrenic nerve into the bathing fluid have been given elsewhere³. Two samples (control and experimental) were collected from each preparation and assayed immediately (cross-over) on leech dorsal muscle, the bathing solution containing equivalent concentration(s) of the experimental drug(s). We also examined the effect of 0.5 μ M atropine on the action potentials of the phrenic nerve.

Oxotremorine sesquifumarate (Oxo-T)—a muscarinic agonist with no post-junctional effect on the diaphragmatic muscle³—induces transmitter release from phrenic nerve endings³ and thus serves as an excellent tool for studying the cholinergic presynaptic mechanisms at this site. The dose-dependent effect of Oxo-T in increasing ACh release and its prevention by preincubation with atropine (Table 1) demonstrate the presence of presynaptic muscarinic receptors on motor nerve terminals. Incubation with 0.5 μ M atropine alone

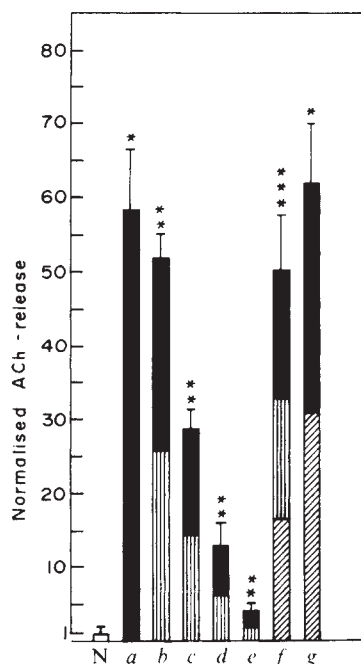


Fig. 1 Histogram showing: a, increase in ACh release caused by 10 μ M Oxo-T sesquifumarate; b, c, d and e, dose-dependent inhibition of a by dopamine hydrochloride (0.005, 0.02, 0.04 and 0.05 μ M, respectively); f, reversal of e in presence of 0.25 μ M pimozide; g, influence of 0.25 μ M pimozide on a. All values are normalised to the mean control ACh release during 15-min periods of electrical stimulation ($N = 4.05$ ng). Values represent means of at least six experiments $\pm$ s.e. One asterisk denotes significant increase from N ($P < 0.05$); two asterisks denote significant reduction from a ($P < 0.05$); three asterisks denote significant increase from e ($P < 0.05$). More than one symbol per bar indicates respective drug interactions. Black indicates Oxo-T; vertical hatching indicates dopamine, and cross hatching indicates pimozide.

Table 1 ACh released by different treatments

Treatment	ACh released (ng)
Control†	4.05 ± 0.73
Atropine sulphate, 0.5 µM	Not measurable (necessarily <1.25)
Oxo-T sesquifumarate, 2.5 µM†	135.14 ± 20.16*
Oxo-T sesquifumarate, 5 µM	208 ± 20.7*
Oxo-T sesquifumarate, 10 µM†	236 ± 33.12*
Atropine sulphate, 0.5 µM plus Oxo-T, 10 µM	Not measurable (necessarily <1.25)
Dopamine hydrochloride, 0.05 µM	4.0 ± 1.1
Dopamine hydrochloride, 5 µM	Not measurable (necessarily <1.25)
Pimozide, 0.25 µM	3.2 ± 0.42

Figures show ACh released, in the form of ACh chloride, into a 4-ml bath from rat phrenic nerve terminals in response to supramaximal stimulation (0.2 ms) at 1 Hz for 15 min in different experimental conditions. Values represent means of at least six experiments ± s.e. Concentrations of drugs are expressed in terms of molarity of the salts.

* Significant difference from control ($P < 0.005$).

† Data taken from ref. 3.

in control experiments inhibited the evoked ACh release to a level below the sensitivity range of our preparation (Table 1).

Dopamine hydrochloride by itself at 0.005–0.05 µM did not affect the normal evoked ACh release (Table 1), but it inhibited, as a function of its concentration, the increased release due to Oxo-T (Fig. 1). In the presence of a greater concentration of dopamine (5 µM), the normal evoked release of ACh was inhibited to a degree that could not be measured (Table 1).

The presence of dopamine receptors with an inhibitory influence on muscarinically (Oxo-T) mediated ACh release was demonstrated by reversal of the dopamine effect in the presence of 0.25 µM pimozide, a specific dopamine-receptor-blocking agent (Fig. 1). Incubation with pimozide alone, unlike atropine, did not affect the normal evoked and Oxo-T-induced release of ACh in control experiments (Table 1 and Fig. 1).

Our results demonstrate for the first time the presence of presynaptic dopaminergic muscarinic receptors on motor nerve endings. A function has been proposed for presynaptic muscarinic receptors at cholinergic sites in the brain⁵ and in the periphery⁶, apparently at variance with our findings. In such studies Oxo-T decreased ACh release and prevented the facilitatory effect of atropine. On the other hand, excess ACh release due to Oxo-T at both central and peripheral sites is well documented^{3,7–9}. Thus it is not clear whether the reported inhibition of ACh release by Oxo-T is mediated through a direct or an indirect mechanism. Indeed incubation with eserine inhibited ACh release from cerebral cortical slices^{10,11}.

The inhibition of normal evoked release of ACh by atropine, at a concentration that failed to affect phrenic nerve action potentials, indicated that the presynaptic muscarinic receptors participate in ACh release from motor nerve endings. Although a similar effect of atropine was reported earlier², its mechanism of action remained obscure. Our study raises the possibility that locally liberated ACh from motor nerve terminals activates the presynaptic muscarinic receptors, leading to further release. The inability of pimozide to increase normal evoked ACh release (Table 1) shows that the inhibitory dopaminergic sites on the motor nerve terminals do not normally influence transmitter release. This is expected because dopamine is not known to be liberated at this site. But in the light of our findings it seems plausible that systemic circulatory dopamine, more so in subjects receiving L-dopa therapy, must modulate ACh release from motor nerve endings, if not at cholinergic synapses in general. This presynaptic dopamine action may be of therapeutic significance because presynaptic cholinergic dominance at the neuromuscular junction and at the junction of motor axon collaterals and Renshaw cells in the spinal cord have been

clearly implicated^{4,12–14} in the 'Oxo-T-model'^{15,16} of parkinsonism. Indeed L-dopa pretreatment was found to protect Oxo-T-induced tremor to some extent^{17,18} and to inhibit the antidromically activated spike discharges of single Renshaw cells¹⁹. Our findings support the hypothesis of presynaptic modulation of ACh release by dopamine at cholinergic sites^{20,21}.

We thank Mr J. R. Vedasiromoni for help in preparing our manuscript.

D. K. GANGULY
M. DAS

*Division of Pharmacology and Experimental Therapeutics,
Indian Institute of Experimental Medicine,
4 Raja S. C. Mullick Road,
Calcutta-700 032, India*

Received 15 November, 1978; accepted 2 March 1979.

- Miyamoto, M. D. & Voile, R. L. *Fedn Proc.* **71**, 1489–1492 (1974).
- Hubbard, J. I. & Wilson, D. F. *Experientia* **26**, 1234–1235 (1970).
- Das, M., Ganguly, D. K. & Vedasiromoni, J. R. *Br. J. Pharmac.* **62**, 195–198 (1978).
- Ganguly, D. K., Nath, D. N., Ross, H.-G. & Vedasiromoni, J. R. *Br. J. Pharmac.* **64**, 47–52 (1978).
- Szerb, J. C. in *Cholinergic Mechanisms and Psychopharmacology* (ed. Jenden, D. J.) 49–60 (Plenum, New York, 1977).
- Kilbinger, H. in *Cholinergic Mechanisms and Psychopharmacology* (ed. Jenden, D. J.) 401–410 (Plenum, New York, 1977).
- Nistri, A. *Naunyn-Schmiedeberg's Archs Pharmac.* **295**, 89–94 (1976).
- Guggenheimer, E. H. & Levinger, I. M. *Experientia*, **31**, 88–89 (1975).
- György, L., Pfeifer, A. K. & Kenyeres, J. *J. Pharm. Pharmac.* **22**, 96–100 (1970).
- Bourdois, P. S., Mitchell, J. F., Somogyi, G. T. & Szerb, J. C. *Br. J. Pharmac.* **52**, 509–517 (1974).
- Szerb, J. C. & Somogyi, G. T. *Nature new Biol.* **241**, 121–122 (1973).
- Ganguly, D. K., Ross, H.-G., Haase, J. & Cleveland, S. *Expl Brain. Res.* **25**, 35–43 (1976).
- Fackler, K., Ross, H.-G., Cleveland, S. & Haase, J. *Pflügers Arch. ges. Physiol.* **368**, R35 (1977).
- Cleveland, S., Ross, H. G., Ganguly, D. K., Kuschmierz, A. & Haase, J. *Pflügers Arch. ges. Physiol.* **373**, R71 (1978).
- Jenden, D. J. in *Selected Pharmacological Testing Methods* Vol. 3 (ed. Burger, A.) 337–361 (Marcel Dekker, New York, 1968).
- Friedman, A. H. & Everett, G. M. *Adv. Pharmac.* **3**, 83–127 (1964).
- Everett, G. M., Mose, P. & Borchering, J. *Fedn Proc.* **30**, 677 (Ab2693) (1971).
- Coward, D. M., Doggett, N. S. & Sayers, A. C. *Arzneimittel-Forsch.* **27**, 2326–2332 (1977).
- Meyer-Lohmann, J., Hellweg, G., Hageneath, R. & Bennecke, R. *Proc. int. Un. Physiol. Sci.* **9**, 389–396 (1972).
- Costa, E. & Cheney, D. L. in *Interactions Between Putative Neurotransmitters in the Brain* (eds Garattini, S., Pujol, J. F. & Samanin, R.) 23–28 (Raven, New York, 1978).
- Vizi, E. S., Ronai, A., Harsingh, L. C. Jr & Knoll, J. in *Cholinergic Mechanisms and Psychopharmacology* (ed. Jenden, D. J.) 587–604 (Plenum, New York, 1977).

Decreased content of immunoreactive enkephalin-like peptide in peripheral tissues of spontaneously hypertensive rats

THE dopaminergic modulation of neuronal transmission in coeliac ganglia of spontaneously hypertensive rats (SHR) has been shown to be less efficient than in control Wistar Kyoto rats (WKY). From these results it was inferred that the development of hypertension may be due to this defect. Recent reports^{2,3} suggest that polypeptides may function as neurotransmitters or neuromodulators in sympathetic ganglia. A heterogeneous group of immunoreactive Met⁵-enkephalin-like (ME) peptides have been found in these ganglia³, and are also present in adrenal glands, in both chromaffin cells and afferent axons (unpublished observations). Some of these peptides, when released from the afferent axons, may modulate opiate receptors located in adrenal medulla (unpublished observations). The vasodilatory action of morphine suggests that if the opiate agonists present in the chromaffin cells were to be secreted they might also cause vasodilation by acting on distant receptors. In view of their possible role in the regulation of sympathetic transmission and vascular tone, we have studied the content and chromatographic characteristics of the immunoreactive ME-like peptides in adrenal gland, sympathetic ganglia, salivary gland and hypothalamus of SHR and WKY rats. We report here

Table 1 The content of ME- and LE-like material in various tissues of WKY and SHR rats

Rat strain	Rat age (weeks)	Coeliac ganglion		Superior cervical ganglion	Salivary gland	Adrenal gland	Hypothalamus
		ME	LE	ME	ME	ME	ME
WKY	4	3.4 ± 0.31 (7)	1.8 ± 0.097 (7)	1.9 ± 0.12 (7)	—	—	—
SHR	4	2.0 ± 0.16* (8)	1.1 ± 0.048* (8)	1.9 ± 0.059 (8)	—	—	—
WKY	8	2.0 ± 0.12 (15)	1.2 ± 0.051 (10)	1.3 ± 0.11 (5)	0.090 ± 0.012 (5)	0.046 ± 0.0081 (5)	3.7 ± 0.14 (10)
SHR	8	0.95 ± 0.028* (15)	0.68 ± 0.030* (10)	0.78 ± 0.20† (5)	0.045 ± 0.010† (5)	0.020 ± 0.004† (5)	3.3 ± 0.17 (10)
WKY	12	1.7 ± 0.14 (5)	—	1.4 ± 0.081 (5)	—	—	—
SHR	12	0.93 ± 0.082† (5)	—	0.93 ± 0.18† (5)	—	—	—

Immediately after dissection, the ganglia were added to boiling 0.1 M CH₃COOH and heated for 3–5 min at 100 °C. The neutralised supernatant was then radioimmunoassayed for ME and LE. Adrenal, salivary glands and hypothalamus were treated by microwaves and extracted with 0.1 M CH₃COOH. Values are expressed as ng per mg protein. No difference was noted in the weights or the protein content of the various tissues from WKY or SHR.

* $P < 0.01$; † $P < 0.05$.

that the content of these peptides was lower in SHR than WKY rats in all tissues studied, except the hypothalamus. In sympathetic ganglia this decrease was due to a reduction in the ME content and not to that of the high molecular weight (MW) immunoreactive ME-like peptides.

Male 3-week-old SHR and WKY rats (Thaconic Farm) were reared in our animal facility in optimal conditions of illumination (12 h light/12 h dark), temperature (24 °C) and humidity. At various ages, an equal number of the SHR and WKY animals were killed by decapitation and exsanguinated. Table 1 shows that SHR already have a lower concentration of immunoreactive ME-like peptides in the coeliac ganglion at 4 weeks after birth, when the blood pressure of WKY and SHR rats is similar¹. Immunoreactive ME-like peptide content is also decreased (Table 1) at 12 weeks in the coeliac and superior cervical ganglia of SHR when their blood pressure is pathologically higher than that of WKY. Also, the coeliac content of Leu⁵-enkephalin-like (LE) polypeptides in SHR is smaller than that in WKY animals. The content of immunoreactive ME-like peptides in the adrenal and salivary glands was also significantly smaller in SHR than in WKY rats (Table 1). In contrast, immunoreactive ME-like peptide content in the hypothalamus was similar in 8-week-old SHR and WKY rats.

The various molecular forms of immunoreactive ME-like peptides present in sympathetic ganglia include a low MW polypeptide which has previously⁵ been shown to be immunologically and chromatographically indistinguishable from ME. We wondered whether in SHR animals there was a

reduced content of ME or of the immunoreactive ME-like high MW polypeptides. Therefore, we extracted the sympathetic ganglia with absolute alcohol which selectively solubilises a polypeptide which on thin layer chromatography behaves like ME, leaving the other high MW immunoreactive ME-like peptides undissolved. We found that a coeliac ganglion of a 12-week-old WKY rat contains about 0.27 ng ME, whereas the homologous ganglion of SHR rats contains only about 0.18 ng ME. A superior cervical ganglion of WKY and SHR contains 0.087 and 0.055 ng ME, respectively. The chromatographic profile of the immunoreactive ME-like peptides present in coeliac ganglia shows that the ganglia of SHR rats contain less of the low MW immunoreactive ME-like peptides but about an equal amount of the high MW immunoreactive material (Fig. 1). This abnormality is present in peripheral tissues but not in the hypothalamus (Table 1) or in the striatum (data not shown). The decrease of peripheral immunoreactive ME-like peptides may be due to a deficit in the enzyme that forms ME, an increase in the enzyme that catabolises ME or an increase in the rate of ME utilisation. Preliminary data from our laboratory^{3,4} indicate that the ratio of high MW immunoreactive ME-like peptides to ME in brain is much smaller than that in peripheral tissues, and we have also shown that the high MW immunoreactive ME-like peptide can release ME when subjected to controlled proteolysis (unpublished observations). Thus, it is possible that the mechanisms which synthesise ME in brain and in periphery differ and that the genetic defect in SHR rats which reduces the content of ME-like peptides in peripheral tissues may be an abnormality in ME biosynthesis. If this defect is an important cause of the hypertension in SHR animals, one could infer that ME-like peptides, like morphine, may lower the vascular tone. However, the nature, mechanism, site and pathogenetic significance of the defect in ganglionic ME-like peptides reported here remains to be investigated.

A. M. DI GIULIO
H.-Y. T. YANG
W. FRATTA
E. COSTA

Laboratory of Preclinical Pharmacology,
National Institute of Mental Health,
Saint Elizabeths Hospital,
Washington, DC 20032

Received 11 December 1978; accepted 14 February 1979.

1. Lutold, B. E., Karoum, F. & Neff, N. H. *Circulation Res.* (in the press).
2. Schultzberg, M. *et al. Neuroscience* 3, 1169–1186 (1978).
3. Di Giulio, A. M. *et al. Neuropharmacology* 17, 989–992 (1978).
4. Costa, E., Di Giulio, A., Fratta, W., Hong, J. & Yang, H.-Y. T. In *Proc. 4th int. Congr. on Catecholamines* (ed. Usdin, E.) (Pergamon, Oxford, in the press).
5. Yang, H.-Y. T., Fratta, W., Hong, J. S., Di Giulio, A. M. & Costa, E. *Neuropharmacology* 17, 433–438 (1978).
6. Eckenhoof, J. E. & Oech, S. R. *Clin. Pharmac. Ther.* 1, 483–524 (1960).
7. Gomes, C., Svensson, T. H. & Trolin, G. *Naunyn-Schmiedeberg's Archs Pharmac.* 294, 141–147 (1976).
8. Yang, H.-Y. T., Hong, J. S. & Costa, E. *Neuropharmacology* 16, 303–307 (1977).

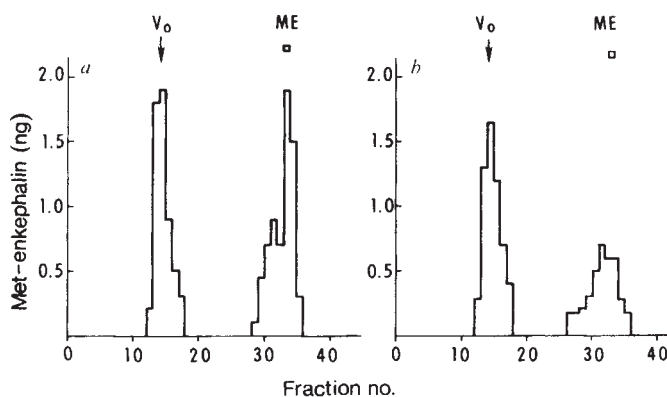


Fig. 1 Biogel P-2 column chromatography of sympathetic ganglia from WKY (a) and SHR (b) rats. Coeliac ganglia (30) and superior cervical ganglia (60) were extracted in 1 M CH₃COOH and heated for 5 min at 100 °C. The resulting supernatant was lyophilised, dissolved in 0.5 ml of 1 M CH₃COOH and chromatographed with Biogel P-2 column (0.9 × 60 cm) using 1 M CH₃COOH as the eluant. Aliquots of each fraction of the eluate were lyophilised and radioimmunoassayed with antiserum directed against ME. Two fractions appear, one, of high MW, is eluted with the void volume (V₀), the other is eluted within the volume that is characteristic of ME. The exclusion limit of this column is 1,800 daltons.

Histamine receptors on rabbit blastocyst and endometrial cell membranes

DESPITE much study, the role of histamine in ovum implantation remains unclear¹⁻⁵. Histamine action is mediated by H_1 and H_2 receptors^{6,7}, which are widely distributed on the cell membranes of mammalian tissues⁸⁻¹⁴. However, no information is available about the histamine receptors on blastocyst or endometrial cell membranes. Recent evidence indicates that a certain lymphocyte subpopulation carries histamine receptors⁹ and has an important regulatory function *in vitro*¹⁰⁻¹² and *in vivo*^{12,13} in mice⁹⁻¹¹, guinea pigs¹² and man¹³. Murine T cells, capable of regulating antibody secretion *in vitro*¹⁰, and precursors of T cells, capable of mediating cytotoxic effector cells¹¹, carry histamine receptors. The density of these receptors increases with cell maturation¹¹. Cell activation by histamine results in increased levels of intracellular cyclic AMP¹¹ and production of small molecular weight material, called histamine suppressor factor, capable of suppressing both cell proliferation and lymphokine production by sensitised lymphocytes following exposure to the antigen¹². Histamine suppression of delayed hypersensitivity in guinea pigs *in vivo* can be blocked completely or partially by H_2 and H_1 antagonists, respectively¹². Also, the administration of H_2 antagonists to man results in enhancement of delayed skin tests¹³. These observations indicate that lymphocytes carrying H_2 receptors acquire immunosuppressor potential on activation.

We demonstrate here that 6-day rabbit blastocyst and uterine endometrial cells carry H_2 and H_1 receptors, respectively. The possible significance of H_1 and H_2 receptors in ovum implantation and in suppressing *in situ* maternal capacity to reject the implanting embryo carrying alloantigens are discussed.

Six-day blastocysts from superovulated¹⁵ domestic rabbits were recovered by flushing the uteri with chilled medium RPMI 1640 supplemented with bovine serum albumin (BSA). The blastocysts were washed twice in the same medium, transferred to a small disposable sterile Petri dish, ruptured and gently teased with a needle. The preparation was passed through a 25-gauge needle to obtain single-cell suspensions. Endometrial cells from the same animals were obtained by scraping off the uterine endometria with a scalpel blade. The scrapings were transferred to a sterile plastic tube containing BSA-supplemented cold RPMI 1640 medium. The clumps of endometrial cells were first passed through a 19-gauge needle and then a few times through a 25-gauge needle to obtain single-cell preparations. The blastocyst and endometrial single-cell suspensions were centrifuged at 720g, washed and resuspended in phosphate-buffered saline (PBS) with 10% fetal calf serum. Cell numbers were then counted and adjusted to 0.5×10^5 cells ml^{-1} . With this technique, less than 1% of cells were found to be dead by the standard Trypan blue exclusion method.

Histamine receptors were visualised by incubating single-cell preparations in fluorescein isothiocyanate-BSA-histamine (FITC-BSA-H) conjugate for 30 min at 4 °C. FITC-BSA-H was prepared by conjugating 50.0 mg FITC-BSA (fluorescein-protein ratio, 3:1) to 350.0 mg histamine dihydrochloride (Calbiochem) in the presence of 1.0 g of 1-ethyl-3(3-dimethylaminopropyl)carbodiimide (ECDI; Story) at room temperature for 2 h followed by extensive dialysis at 4 °C against 2 l of PBS for 48 h. In all the experiments, FITC-BSA-H was used at a dilution of 1:25 with PBS. Following incubation the cells were washed twice, mounted in PBS-glycerine mixture and examined under a fluorescence microscope using an epillumination system. Membrane stain was demonstrated as fluorescent speckles. The control cells, incubated in FITC-BSA in the absence of histamine, were negative. Moreover, red blood cells, as contaminants in the endometrial cell suspensions, did not

show any binding with FITC-BSA-H conjugate. The specificity of binding was determined by incubating cell suspensions in free histamine, chlorpheniramine (H_1 -receptor antagonist)¹⁶ and metiamide (H_2 -receptor antagonist)¹⁷ before incubation in FITC-BSA-H. Statistical analysis was according to Student's *t* test.

Figure 1a shows the results of histamine binding with blastocyst cell membranes: $37 \pm 11\%$ of blastocyst cells showed bright membrane fluorescence on staining with FITC-BSA-H. Histamine and metiamide, at both 10^{-4} M and 10^{-5} M concentrations, displaced the binding significantly ($P < 0.01$). Chlorpheniramine significantly displaced the binding only at 10^{-4} M ($P < 0.05$). Figure 1b shows the results of histamine receptor determinations on endometrial cell membranes: $21 \pm 9\%$ of endometrial cells showed binding with FITC-BSA-H conjugate and this binding was significantly suppressed by free histamine and chlorpheniramine at both 10^{-4} M ($P < 0.005$) and 10^{-5} M ($P < 0.05$). Metiamide, on the other hand, did not suppress the binding. None of the above agents at 10^{-6} M were able to alter the binding response of either blastocyst or endometrial cell membranes. The above results indicate that blastocyst cell membranes carry mainly H_2 receptors, whereas endometrial cells carry H_1 receptors.

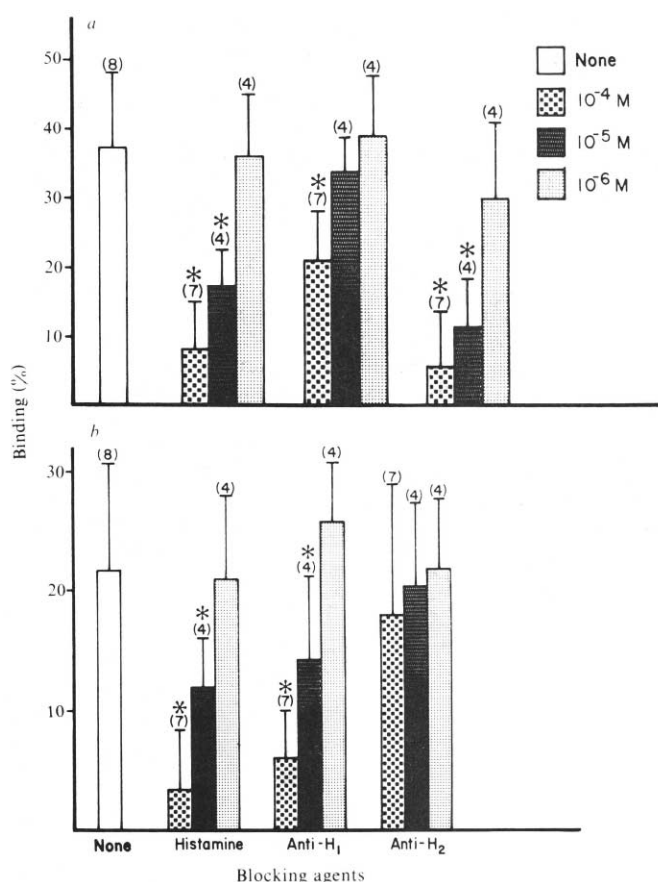


Fig. 1 The blocking effects of histamine or various histamine antagonists on the binding of blastocyst (a) and endometrial (b) cell suspensions to FITC-BSA-H. Results shown are means $\pm$ s.d. of several experiments. The number of experiments is shown in parentheses. Asterisks denote a significant difference from binding values in the absence of a blocking agent. Rabbit blastocyst or endometrial cell suspensions (0.5×10^5 in 1.0 ml of RPMI 1640) were incubated for 30 min at 4 °C with 0.1 ml of various concentrations of the blocking agents, histamine, anti H_1 (chlorpheniramine) or anti- H_2 (metiamide). They were washed three times in chilled medium, stained by incubating them with FITC-BSA-H at 4 °C for 30 min and then washed and mounted. In all experiments at least 200 cells per slide were counted.

The mammalian blastocyst is composed of two types of cell: the inner cell mass and trophoblast. Similarly, the endometrium is composed of epithelial, stromal and glandular cells. Our findings do not indicate the cell type in the blastocyst or endometrium that carries histamine receptors. The observation that histamine receptors could be detected on some of the blastocyst or endometrial cells is not unique; human blood lymphocytes are also heterogeneous and only $30 \pm 3.8\%$ of these cells can bind to histamine¹⁸.

Further work is necessary to elucidate the possible role of histamine receptors in blastocyst-endometrium metabolism, implantation and acceptance of the blastocyst allograft by the maternal tissues¹⁹. Despite many suggestions, it remains a challenge to reproductive biologists and immunobiologists to explain the escape of the embryo from the immunological responses of the mother¹⁹. In an immunologically competent system, histamine binding to H₁ receptors mediates a variety of pro-inflammatory events, such as increased permeability and vasodilatation, whereas histamine activation of H₂ receptors mediates several immunosuppressive effects through stimulation of cyclic AMP synthesis¹¹. A local inflammatory reaction, accompanied by accumulation of histamine in the uterus, occurs around the time of implantation²⁰. Implantation in the rabbit begins within 6½–7 d of pregnancy (day 1 = 24 h post coitum)²¹, and the cyclic-AMP/cyclic GMP ratio in the rabbit blastocyst changes before implantation²². Moreover, cyclic AMP has been shown to participate in blastocyst implantation in the mouse^{23,24}. The above bimodal function of histamine through H₁ and H₂ receptors is feasible, as systemic administration of either H₁ or H₂ receptor blockers in rats has no effect on preimplantation stages of pregnancy, whereas the combination of the two interferes with implantation²⁵. Furthermore, histamine involvement in inducing implantation in mice or rats during delayed implantation^{26,27}, and interruption of implantation in rabbits by inhibition of uterine histamine release have been reported recently²⁸. Also, 6-day rabbit blastocysts have been shown to suppress the mitogen response of autologous splenic lymphocytes *in vitro*²⁹. It remains to be seen whether this suppressor activity of blastocysts is mediated through activation of histamine receptors.

We thank Mrs Marilyn Ousley and Mr Terrill Smith for technical assistance, Mrs Anne Scott for typing the manuscript, and Dr John G. Paul (Smith, Klein and French) for supply of chlorpheniramine and metiamide. This research was supported in part by NIH Biomedical Research Support grants 5 S01-RR-5373 and HD-12304, and Mid America Cancer Center Program.

S. K. DEY
C. VILLANUEVA
N. I. ABDOU

Departments of Obstetrics and Gynecology and Medicine,
Ralph L. Smith Human Development Research Center,
University of Kansas Medical Center,
Kansas City, Kansas 66103

Received 22 November 1978; accepted 19 February 1979.

- Shellesnyak, M. C. *Am. J. Physiol.* **170**, 522–527 (1952); *Recent Prog. Horm. Res.* **13**, 269–322 (1957).
- Marcus, G. J., Kracier, P. F. & Shellesnyak, M. C. *J. Reprod. Fert.* **5**, 409–415 (1963).
- Marcus, G. J., Shellesnyak, M. C. & Kracier, P. F. *Acta endocr.* **47**, 255–264 (1964).
- Nalbandov, A. V. in *The Biology of the Blastocyst* (Blandau, R. J. ed.) 386–392 (University of Chicago Press, 1971).
- DeFeo, V. J. in *Cellular Biology of the Uterus* (Wynn, R. M. ed.) 191–290 (Appleton-Century-Crofts, New York, 1967).
- Ash, A. S. F. & Schild, H. O. *Br. J. Pharmac. Chemother.* **27**, 427–439 (1966).
- Black, J. W., Duncan, W. A. M., Durant, G. J., Ganellin, C. R. & Parsons, M. *Nature* **236**, 385–390 (1972).
- Onaya, T., Hashizume, K., Sato, A., Takazawa, T., Akasur, F. & Endo, W. *Endocrinology* **100**, 61–66 (1977).
- Weinstein, Y. & Melmon, K. L. *Immun. Commun.* **5**, 401–416 (1976).
- Shearer, G. M., Simpson, E., Weinstein, Y. & Melmon, K. L. *J. Immun.* **118**, 756–761 (1977).
- Roszkowski, W., Plaut, M. & Lichtenstein, L. M. *Science* **195**, 683–685 (1977).
- Rocklin, R. E. *J. Immun.* **118**, 1734–1738 (1977); *J. clin. Invest.* **57**, 1051–1058 (1976).
- Avella, J., Binder, H. J., Madsen, J. E. & Askenase, P. W. *Lancet* **i**, 624–626 (1978).

- Ray Wang, S. & Zweiman, B. *Cell. Immun.* **36**, 28–36 (1978).
- Mukherjee, A., Dey, S. K., Sen Gupta, J., Ramadoss, C. R. & Dickmann, Z. *J. Reprod. Fert.* **53**, 77–80 (1978).
- Levi, R. & Capurro, N. in *Int. Symp. Histamine H₂-Receptor Antagonist* (eds Wood, D. J. & Simkins, M. A.) 175–186 (Smith, Kline and French Laboratories, Welwyn Garden City, 1973).
- Black, J. W. *et al. Ag. Actions* **3**, 133–137 (1973).
- Ousley, M., Sagawa, A. & Abdou, N. I. *Fedn Proc.* **36**, 1244 (1977).
- Amoroso, E. C. & Perry, J. S. *Phil. trans. R. Soc. B* **271**, 343–361 (1975).
- Marcus, G. J. & Shellesnyak, M. C. *Acta endocr.* **57**, 136–141 (1968).
- Dey, S. K., Dickmann, Z. & Sen Gupta, J. *Steroids* **28**, 481–485 (1976).
- Dey, S. K., Kimura, F., Mukherjee, A. & Dickmann, Z. *J. Reprod. Fert.* **52**, 235–237 (1978).
- Holmes, P. V. & Bergstrom, S. J. *Reprod. Fert.* **43**, 329–332 (1975); *Am. J. Obstet. Gynec.* **124**, 301–306 (1976).
- Webb, F. T. G. *J. Reprod. Fert.* **42**, 511–517 (1975).
- Brandon, J. M. & Wallis, R. M. *J. Reprod. Fert.* **50**, 251–254 (1977).
- Brandon, J. M. A. *Ming Soc. Study of Fertility*, Abstr. no. 13 (1977).
- Saksena, S. K., Lau, I. F. & Chang, M. C. *Acta endocr.* **81**, 801–807 (1976).
- Dey, S. K., Villanueva, C., Chien, S. M. & Crist, R. D. *J. Reprod. Fert.* **53**, 23–26 (1978).
- Dey, S. K., Villanueva, C. & Abdou, N. I. *Fedn Proc.* **37**, 1474 (1978).

Early pregnancy factor is immunosuppressive

IMMUNOSUPPRESSION during pregnancy has important consequences not only with respect to the recognition of the antigenically dissimilar embryo¹ but also in terms of susceptibility to viral or tumour assaults^{2,3}. We have recently described an early pregnancy factor (EPF) detectable within 24 h of mating and have postulated that this protein is responsible for depression of lymphocyte activity during pregnancy^{4–6}. Immunosuppressive properties have been attributed to other proteins detectable during pregnancy. These include pregnancy-associated macroglobulin^{7,8}, α -fetoprotein⁹ and human chorionic gonadotropin (HCG)^{10,11}. One characteristic these substances have in common, but in which they differ from EPF, is that they are not present, or not detectable, in the early stages of pregnancy. Recently the immunosuppressive activity of HCG has been disputed by several workers working with highly purified material^{12,13}. We have now shown that EPF inhibits the expression of an immune response *in vivo*.

After fertile matings, EPF has been detected on the lymphocytes and in the serum of mice, humans and sheep within 6–24 h and has been shown to be present for approximately the first two trimesters of pregnancy^{4–6}; EPF activity was not detected after sterile matings in any of the three species (refs 4, 6, B. Fitzgerald, personal communication). This substance was found in serum in two different forms, one of apparent molecular weight 200,000–240,000 detectable from 6 h after mating, and the other, ~50,000 MW detectable from 6 d after mating¹⁴.

The rosette inhibition test has been used as the assay method for the detection of EPF. This test was originally used to assess the immunosuppressive activity of anti-lymphocyte serum (ALS)^{15,16} and the results showed that the rosette inhibition titre (RIT) of each serum did not vary significantly using lymphocytes from different normal donors. If the lymphocytes were from pregnant donors, or had been incubated in serum from pregnant donors, the RIT of the ALS was significantly increased^{4–6}. This was a novel use of the rosette inhibition test and, on the basis of the ALS-sparing properties of EPF, we proposed that EPF is immunosuppressive. The use of an *in vivo* test, however, would clearly establish this point. We have now used the technique of adoptive transfer of contact sensitivity to trinitrochlorobenzene (TNCB) by lymphoid cells from sensitised mice^{17,18}. Lymph node cells, taken 4 d after sensitising mice with TNCB, were incubated with EPF preparations from pregnant sheep (Table 1), before injection of these cells into naive syngeneic mice. Skin tests for delayed-type hypersensitivity were immediately performed on the recipients¹⁹.

Serum or serum fractions having EPF activity (Table 1) suppressed the adoptive transfer of contact sensitivity (Table 2). When TNCB-sensitised lymph node cells were incubated with 24-h pregnant sheep serum (EPF I, Table 1), the ability of these

cells to transfer TNCB sensitivity was completely abolished (Table 2, experiment 1). Partial suppression of activity was observed when cells were preincubated with normal sheep serum, but significant transfer of activity was still achieved. Preincubation with the high MW fraction (EPF II, Table 1) also abolished transfer of TNCB sensitivity, whereas the use of the corresponding non-pregnant sheep serum fraction (control II) had no effect (Table 2, experiment 2). Two preparations of the low MW fraction prepared from 24-h pregnant sheep serum (EPF IIIa and EPF IIIb, Table 1) completely suppressed adoptive transfer, compared with normal serum controls. Although one of these controls (control IIIa) gave a slight decrease in reaction, there was still a significant transfer of sensitivity. Control IIIb gave a result not significantly different from that obtained with the cells incubated in medium alone (Table 2, experiment 3). A preparation of the low MW fraction obtained from the serum of 1–2-month pregnant sheep (EPF IV, Table 1) completely abolished adoptive transfer, whereas the corresponding fraction from normal serum (control IV) had no significant effect (Table 2, experiment 4).

In these experiments we have shown that EPF inhibits the expression of an immune response *in vivo*; the suppressive effect demonstrated was not species specific as sheep EPF inhibited the activity of mouse lymphocytes. Previously EPF activity had been detected using the rosette inhibition test and, although this test had proved convenient and very accurate, its role as a measure of immunological function was still debatable. The demonstration that EPF can completely suppress the adoptive transfer of contact sensitivity, well established as an immunological process, clearly indicates that EPF is immunosuppressive *in vivo*. The indirect technique of adoptive transfer was used, as preliminary experiments involving the injection of sheep EPF directly into mice, both before and after sensitisation, indicated

Table 2 Suppression of adoptive transfer of contact sensitivity by EPF

Expt	Cells transferred*	Cells preincubated with†	No. of recipient mice	Skin test‡ 24 h ear swelling ±s.d. (cm × 10 ⁻³)
1	None	–	6	4.3 ± 1.0
	80 × 10 ⁶	–	5	9.4 ± 3.5
	80 × 10 ⁶	Control I	5	6.1 ± 1.4
	80 × 10 ⁶	EPF I	5	4.1 ± 2.3
2	None	–	6	3.2 ± 1.3
	160 × 10 ⁶	–	3	7.3 ± 2.8
	160 × 10 ⁶	Control II	6	7.9 ± 2.6
	160 × 10 ⁶	EPF II	6	4.5 ± 1.8
3a	None	–	5	4.1 ± 1.7
	100 × 10 ⁶	–	5	8.3 ± 2.0
	100 × 10 ⁶	Control IIIa	5	6.6 ± 1.4
	100 × 10 ⁶	EPF IIIa	5	4.1 ± 1.8
3b	None	–	4	5.4 ± 1.6
	130 × 10 ⁶	–	5	8.9 ± 3.4
	130 × 10 ⁶	Control IIIb	5	9.6 ± 2.6
	130 × 10 ⁶	EPF IIIb	5	5.8 ± 1.9
4	None	–	6	4.3 ± 1.0
	80 × 10 ⁶	–	5	9.4 ± 3.5
	80 × 10 ⁶	Control IV	4	7.5 ± 2.4
	80 × 10 ⁶	EPF IV	5	4.6 ± 1.6

* Axillary and inguinal lymph node cells taken from CBA mice 4 d after skin painting with 7 mg TNCB in 0.1 ml acetone¹⁹ were washed three times in culture medium and preincubated at 50 × 10⁶ cells ml⁻¹ in reaction mixture as indicated.

† Experiment 1: cells were incubated at 37 °C for 60 min in Eagle's basal medium containing 10% fetal calf serum (Commonwealth Serum Labs) and 50% control I or EPF I (Table 1). Cells were washed twice in culture medium, suspended in saline and injected intraperitoneally (0.2 ml) into the recipients (CBA mice). Experiment 2: as above, but reaction mixture contained 20% of control II or EPF II (Table 1). Experiments 3a and 3b: as above, but reaction mixture contained 20% of control IIIa, EPF IIIa, control IIIb or EPF IIIb (Table 1). Experiment 4: as above, but reaction mixture contained 20% of control IV or EPF IV (Table 1).

‡ Within 1 h of cell transfer, both ears of control and recipient mice were measured using an engineer's micrometer (Mitutoyo) and painted with 10 μl of a 1% solution of TNCB in acetone. After 24 h the ear thickness was again measured and ear swelling determined by subtraction.

Table 1 Preparation of serum and serum fractions containing EPF

Source	EPF I Whole serum (24 h)	EPF II Serum fraction (24 h)	EPF III DEAE- Sephadex treated serum (24 h)	EPF IV Serum fraction (1–2 mth)
Approximate MW	240,000	240,000	50,000	50,000
*Rosette inhibition titre	28	24	(a)24 (b)22	26

EPF I, pregnant sheep serum, taken 24 h after mating, was inactivated at 56 °C for 30 min, and the rosette inhibition titre (RIT) determined by the method described previously¹⁶. Serum taken from a nonpregnant ewe was similarly treated (control I). EPF II, pregnant sheep serum, as above, was fractionated on an Ultrogel AcA34 (LKB) column¹⁴ and tested for EPF activity by the RIT. The fractions that contained the EPF activity in the region of MW 240,000 were pooled, dialysed against Hank's balanced salt solution (HBSS) and concentrated to 1/5th of the original serum volume. The RIT of the preparation was determined before use. Nonpregnant ewe serum was similarly processed (control II). EPF IIIa and EPF IIIb, serum was collected from two pregnant ewes, 24 h after mating. Thirty ml of each were partially purified with DEAE-Sephadex A50 (Pharmacia), which removed the macroglobulin and albumin fractions of the serum proteins, then dialysed against HBSS and concentrated to 1/5th of the original serum volume. Before use, the RIT of these preparations was estimated and they were fractionated to determine the MW of the EPF activity¹⁴. Serum from two nonpregnant ewes was similarly treated (control IIIa and IIIb). EPF IV, serum was collected from pregnant sheep 1–2 months after mating and applied to an Ultrogel column. The fractions, MW ~50,000, which had EPF activity, were pooled, dialysed against HBSS and concentrated to 1/5th of the original serum volume. Control IV was nonpregnant ewe serum run in parallel.

* The rosette inhibition titre (RIT) is expressed as the logarithm (base 2) of the reciprocal × 10⁻³ of the highest dilution of anti-lymphocyte serum (ALS) to inhibit the rosette formation to 75% or less than that of control test (cells in HBSS alone). RIT obtained using serum from non-pregnant sheep: mean = 9.8, s.e.m. = 0.3, n = 20.

that immunosuppression was partially obscured by the effect of heterologous protein.

Adoptive transfer of contact sensitivity and inhibition of rosette formation are both T-cell dependent^{16,17}. As EPF affects both of these reactions, its primary function in protecting the fetus may be directed towards a depression of T-cell activity in the maternal system. We may speculate that EPF might protect the mother from immunological attack by the fetus. EPF has been detected in the fetal tissues (H.M. *et al.*, unpublished observations), which suggests another possible role for EPF, namely in the regulation of the development of the fetal immune system. The mechanism of action of EPF is still far from clear; its action is obviously not antigen specific and not even species specific. Whether it is suppressive by simply coating lymphocyte surfaces and preventing antigen presentation, or by a more active method, is yet to be established.

This work was supported by the National Health and Medical Research Council and the Queensland Cancer Fund.

F. P. NOONAN
W. J. HALLIDAY

Department of Microbiology,
University of Queensland

H. MORTON*
G. J. A. CLUNIE

Department of Surgery,
Princess Alexandra Hospital,
University of Queensland,
Brisbane, Queensland, Australia

Received 11 September 1978; accepted 27 February 1979.

* To whom correspondence should be addressed at: Department of Surgery, Princess Alexandra Hospital, Ipswich Road, Woolloongabba, Brisbane, Queensland, Australia 4102.

1. Kerr, M. G. *J. Reprod. Fert. Suppl.* 3, 49–55 (1968).
2. Scott, J. S. *J. Reprod. Fert. Suppl.* 3, 41–48 (1968).
3. Paccs, A. S. & Pettsik, B. *Lancet* i, 330–331 (1977).
4. Morton, H., Hegh, V. & Clunie, G. J. A. *Proc. R. Soc. B* 193, 413–419 (1976).

5. Morton, H., Rolfe, B., Clunie, G. J. A., Anderson, M. J. & Morrison, J. *Lancet* **1**, 394–397 (1977).
6. Morton, H., Nancarrow, C. D., Scaramuzzi, R. J., Evison, B. M. & Clunie, G. J. A. *J. Reprod. Fert.* **56**, (in the press).
7. Stimson, W. H. & Eubank-scott, L. *FEBS Lett.* **23**, 298–302 (1972).
8. von Schoultz, B. & Stigbrand, T. *Acta obstet. gynec. scand.* **52**, 51–57 (1973).
9. Yachnin, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2857–2861 (1976).
10. Contractor, S. F. & Davies, H. *Nature new Biol.* **243**, 284–286 (1973).
11. Adcock, E. W. *et al. Science* **181**, 845–847 (1973).
12. Caldwell, J. L., Stites, D. P. & Fudenberg, H. H. *J. Immun.* **115**, 1249–1253 (1975).
13. Gundert, D., Merz, W. E., Hilgenfeldt, U. & Brossmer, R. *FEBS Lett* **53**, 309–312 (1975).
14. Clarke, F. M., Morton, H. & Clunie, G. J. A. *Clin. exp. Immun.* **32**, 318–323, (1978).
15. Bach, J. F., Dormont, J., Dardenne, M. & Balner, H. *Transplantation* **8**, 265–280 (1969).
16. Morton, H., Hegh, V. & Clunie, G. J. A. *Aust. J. exp. Biol. med. Sci.* **53**, 373–380 (1975).
17. Asherson, G. L. & Zembala, N. *Nature new Biol.* **244**, 176–177 (1973).
18. Moorhead, J. W. *J. Immun.* **119**, 1773–1777 (1977).
19. Noonan, F. & Halliday, W. J. *Int. Archs Allergy appl. Immun.* **56**, 523–532 (1978).

Myosin as a constituent of the *Drosophila* egg cortex

MYOSIN and actin have been found in many cell types, including the eggs of *Amphibia*¹, sea urchins², starfish³ and rats⁴. Their distribution correlates well with the distribution of microfilaments in structures such as the contractile ring, and these proteins are believed to have a major role in cell cleavage and plasmalemma elongation⁵. They have not previously been recorded in *Drosophila* eggs and embryos. The egg of *Drosophila* develops as a syncytium with about 6,000 nuclei present in a common cytoplasm⁶. Subsequently, nearly all the nuclei are simultaneously surrounded by plasmalemmas which grow down into the cortex of the egg. Because of this, and because of the existence of several mutants which specifically affect cellularisation^{7,8}, *Drosophila* eggs are a particularly appropriate system in which to study cell membrane formation. The simultaneous

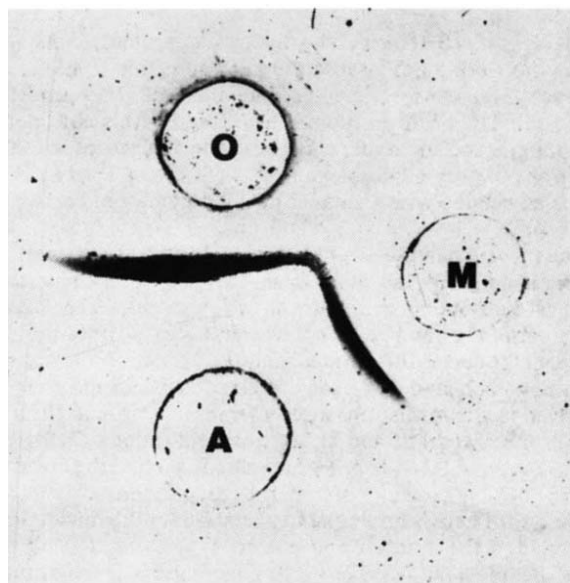


Fig. 1 Precipitin bands against 20 µg fly flight muscle in 2% SDS and 60 mM Tris-HCl, pH 6.8 (M), and 80 µg of an extract of locust oocytes (O). The extract was obtained by homogenising in 1.2 M KCl, 15 mM Na pyrophosphate and 15 mM Tris-HCl, pH 7. The 50% (NH₄)₂SO₄ cut was dissolved in 0.6 M KCl, 2 mM dithiothreitol (DTT), 2 mM ATP and 15 mM Tris-HCl and dialysed against 0.25 M KCl, 2 mM DTT, 2 mM ATP and 15 mM Tris-HCl, pH 7. The precipitate was dissolved in 2% SDS and 60 mM Tris-HCl, pH 6.8. Ouchterlony plates contained 2% Noble's agar, 0.5 M NaCl, 0.01 M phosphate buffer, pH 7.5, 0.1% SDS and 0.5% Triton and were left for 4 d at 4 °C before washing for 2 d in 0.6 M KCl and 20 mM Na pyrophosphate, pH 7, 1 d in 0.8% NaCl, followed by drying down and staining. 'A' marks the antiserum well.

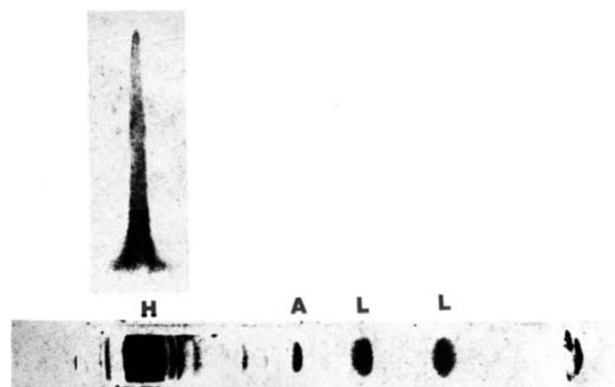


Fig. 2 Crossed immunoelectrophoresis to demonstrate antibodies against myosin. *Sarcophaga* flight muscle myosin (14 µg) was electrophoresed on a tube gel with SDS by the method of Weber and Osborn¹⁴. Electrophoresis in the second dimension was by the method of Converse and Papermaster¹⁵. A strip (1 mm thick) from the SDS gel was electrophoresed into 1% agarose; the layer of agarose (0.5 cm wide) next to the SDS gel contained 0.5% Tween and the rest of the agarose contained 2% IgG from antiserum. The stained SDS gel is heavily loaded to show minor bands. H indicates the heavy chain of myosin, L the light chains, and A contaminating actin.

cellularisation of the whole of the egg cortex would seem to require the existence of large numbers of microfilaments but these have not yet been visualised in the electron microscope⁹. We report here the finding of a continuous band of myosin immediately below the egg plasmalemma which could well be associated with massed microfilament bundles. This is laid down by the time of fertilisation and remains continuous until the arrival of the nuclei in the cortex. The structure of the band and its relationship to early developmental events are described.

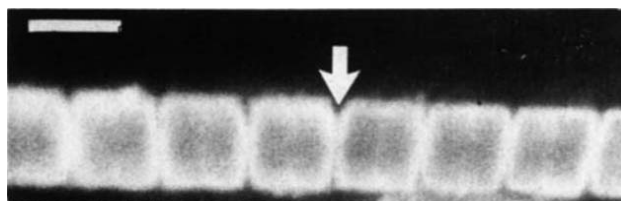


Fig. 3 Freshly dissected *Drosophila* flight muscle myofibril decorated with anti-myosin at a final concentration of 2 mg ml⁻¹. Decoration technique as for embryo sections. Arrow indicates Z line. Scale bar, 3 µm.

Antibodies were made against *Sarcophaga* (blowfly) flight muscle myosin. Myosin was prepared by the method of Bullard and Reedy¹⁰, and further purified by running on 5% SDS-acrylamide slab gels. The band containing the heavy chain of myosin was identified, cut out, homogenised with four volumes of phosphate-buffered saline (PBS) plus 0.5 ml Freund's adjuvant, and injected into white New Zealand rabbits. The total volume was about 3 ml. A total of 1 mg of myosin heavy chain was injected in three equal fortnightly injections. The rabbit was bled 1 week after the last injection. IgG was purified from the antiserum following the method of Campbell *et al.*¹¹. This antiserum gave a single band when tested in Ouchterlony plates (Fig. 1). It also reacted with an extract of locust oocytes containing a putative myosin as a major component. The fusion of the precipitin arcs formed with muscle myosin and oocyte extract shows that the antigens in these preparations had common antigenic sites. This situation is like that of the starfish, where antiserum against egg myosin reacts with tube foot muscle myosin¹². The specificity of the antiserum was further tested in crossed immunoelectrophoresis. A large rocket was obtained

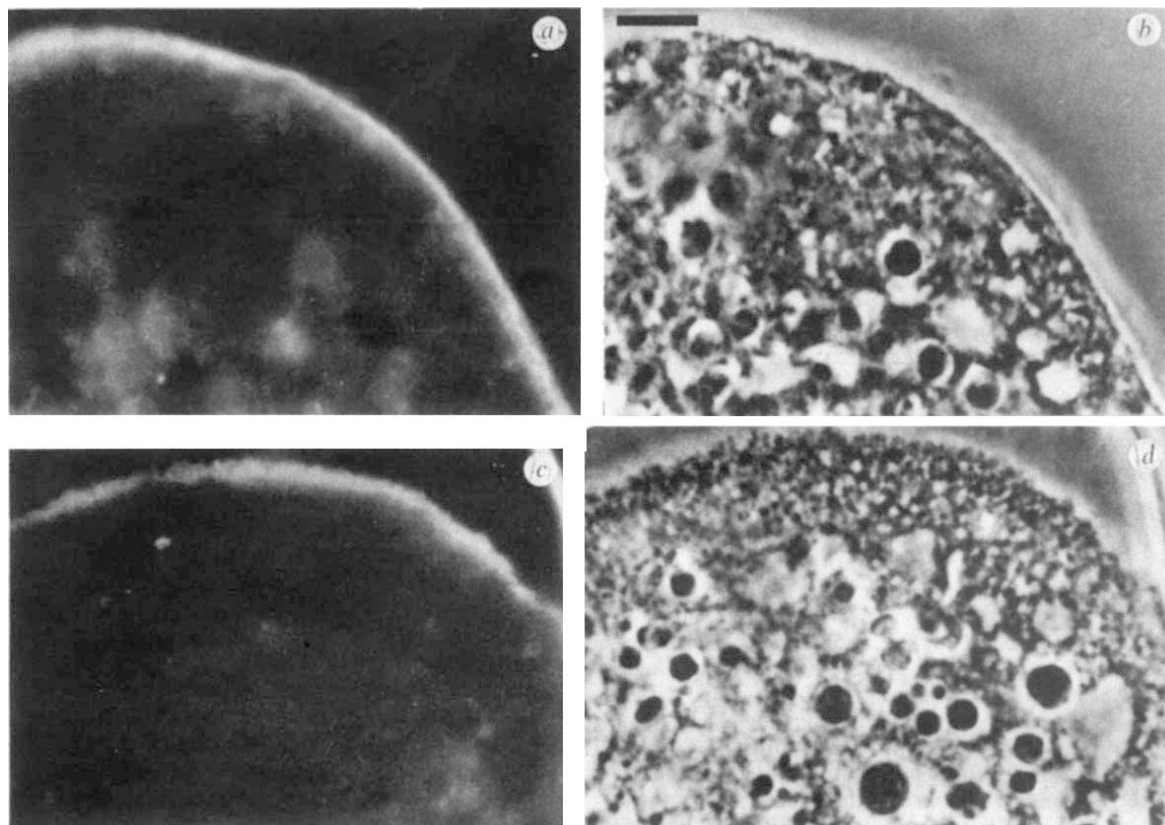


Fig. 4 Sections of 1 nucleus stage *Drosophila* embryos decorated with anti-myosin IgG. In both cases the same section is seen under epifluorescence (a, c) and phase contrast (b, d). a And b show the sidewall running obliquely into the anterior tip region, c and d show the posterior pole. In c and d the brightly fluorescent structure at upper right is the autofluorescent vitelline membrane. Scale bar, 5 μm .

opposite the myosin heavy chain (Fig. 2). Our antiserum binds to A-bands of *Drosophila* flight muscle myofibrils giving patterns similar to those described by Garrett for *Lethocerus* flight muscle myofibrils¹³ (Fig. 3).

Sections of embryos were prepared by a modification of Sainte-Marie's method¹⁶. All procedures except for embedding were carried out at 4 °C. Fixing was done for 1 h using a 6 : 16 : 22 mixture of formaldehyde : ethanol : distilled water. The embryos were carefully punctured to allow adequate penetration of the fixative. They were then dehydrated by passing through each of the following solutions for 30 min: 70% ethanol, 90% ethanol, absolute ethanol and isopropanol (two changes). Clearing was accomplished using ligroin (BDH) for 1 h. The embryos were embedded in paraplast and sectioned at 5 μm .

The technique very effectively preserves delicate structural features such as the cleavage furrows. Sections were treated with either anti-myosin IgG or IgG from the same rabbit before immunisation, using a final concentration of 2 mg ml⁻¹ in buffered saline in both cases. They were incubated at 37 °C for 1 h and then washed for 15 min with PBS, pH 9.5. Results were complicated by nonspecific fluorescence for some pre-immune sera. This was reduced by using a pH 9.5 wash as recommended by Bennett *et al.*¹⁷. Fluorescein isothiocyanate-conjugated sheep anti-rabbit IgG (Wellcome) was added as the second part of the sandwich at a concentration of 0.8 mg ml⁻¹ in PBS and incubated and washed as before. Sections were mounted in a 9 : 1 mixture of glycerol : 1 M Trizma base, pH 9. Microscopy was carried out with a Zeiss standard R microscope equipped with an epifluorescence attachment and a Planapo lens ($\times 63$ NA 1.4).

The results show the existence of a continuous band of fluorescence immediately below the cell membrane, extending around the whole egg (Fig. 4). Controls show no such band, only background fluorescence. Below the band more diffuse material extends deeper into the cortex. Analysis of photographs from 18 embryos (1 nucleus to 32 nuclei stages), indicated a thickness of

1.00 μm (s.d. ± 0.15 μm). This figure is apparently constant all along the sidewall. Towards the posterior pole (Fig. 4c, d) it gradually thickens to give a maximum of 1.59 μm (s.d. ± 0.15 μm). The posterior pole is associated with a considerable development of microvilli, which also show fluorescence. At the anterior pole the band thickens to 1.5 μm (s.d. ± 0.11 μm) but this is associated with a smaller development of microvilli (Fig. 4a, b).

The myosin band seems to be unlike the whorl-like masses of microfilaments in the amphibian cortex¹, but similar to the cortical distribution of myosin in rat oocytes⁴. What may the functions of the band be? We have evidence that once the nuclei arrive in the cortex this band of material breaks up and parts of it becomes associated with each nucleus. It subsequently becomes involved in the formation of cell membranes around the nuclei (R. W., R. Magrath and B. B., in preparation). It therefore seems reasonable to propose that one function of the epicortical myosin band is as a store of materials required in cell membrane formation. There is no simple explanation for the thickenings of the band at the anterior and posterior poles. At the posterior pole, the thickening may be partly concerned with the formation of a double layer of cells, the pole cells and the underlying somatic blastoderm cells. However, this does not explain the increased thickness at the anterior pole.

We thank Alba Warn, Adrian Thompson, Ruth Magrath and Joanna Bell for technical help.

RICHARD WARN

School of Biology,
University of East Anglia

BELINDA BULLARD

Agricultural Research Council Unit,
Department of Zoology,
South Parks Road,
Oxford, UK

School of Biology,
University of East Anglia,
Norwich, Norfolk, UK

SHOHREH MALEKI

Received 19 June 1978; accepted 5 March 1979.

1. Franke, W. *et al. Cytobiologie* **14**, 111–130 (1976).
2. Mabuchi, I. *J. Cell Biol.* **59**, 542–547 (1973).
3. Mabuchi, I. *J. Biochem.* **76**, 47–55 (1974).
4. Amsterdam, A., Lindner, H. R. & Groschel-Stewart, T. *Anat. Rec.* **187**, 311–328 (1977).
5. Shroeder, T. E. in *Molecules and Movement* (eds Inoué, S. & Stephens, R. E.) 305–332 (Raven, New York, 1975).
6. Zalokar, M. & Erk, I. *J. Micro. Biol. Cell* **25**, 97–106 (1976).
7. Zalokar, M., Audit, C. & Erk, I. *Dev. Biol.* **47**, 419–432 (1975).
8. Rice, T. B. & Garen, A. *Dev. Biol.* **43**, 277–286 (1975).
9. Fullilove, S. L. & Jacobson, A. G. *Dev. Biol.* **26**, 560–577 (1977).
10. Bullard, B. & Reedy, M. K. *Cold Spring Harb. Symp. quant. Biol.* **37**, 423–428 (1972).
11. Campbell, D. H., Garvey, J. S., Cremer, N. E. & Sussdorf, D. H. in *Methods in Immunology* 2nd edn, 189–191 (1970).
12. Mabuchi, I. & Okuno, M. *J. Cell Biol.* **74**, 251–263 (1977).
13. Garrett, W. E. thesis, Duke Univ. (1976).
14. Weber, K. & Osborn, M. *J. Biol. Chem.* **244**, 4406–4412 (1969).
15. Converse, C. A. & Papernaster, D. S. *Science* **189**, 469–472 (1975).
16. Sainte-Marie, G. *J. Histochem. Cytochem.* **10**, 250–256 (1962).
17. Bennett, G. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4364–4368 (1978).

The topography of the purple membrane

THE purple membrane of *Halobacterium halobium* functions as a light-driven proton pump^{1,2}. It is about 45 Å thick³ and contains a single polypeptide chain comprising about 240 amino acids⁴. Retinal is covalently attached to a specific lysine residue⁴ and has been implicated in the absorption of incident light energy². The retinal lies at an angle of about 20° to the plane of

the bilayer⁵. Seven protein rods running through the membrane, probably α -helices, have been detected by electron diffraction^{6,7}. Circular dichroism studies indicate a 75% helical content which is consistent with a seven-helix model⁸. A structure containing seven such α -helices requires about 210 amino acids to cross a membrane 45 Å thick seven times, thus implying that most of the polypeptide chain is buried in the lipid bilayer. Digestion of the membrane with proteolytic enzymes supports this view and only three parts of the polypeptide chain are apparently exposed. Here, we show by sequence analysis of digestion products that a section corresponding to the N-terminal seven residues is removed by pronase and proteinase K.

In common with pepsin and papain⁹, pronase and proteinase K also split at several internal sites¹⁰ in the region of residues 71–76 in the sequence of the polypeptide chain (Fig. 1), thereby removing a small segment. A third digestion site for these enzymes is in the region of the C-terminus. Trypsin, which does not split the chain internally, liberates into solution a 21-residue peptide from the C-terminus (residues –1 to –21 in Fig. 1) but does not split the bond between residues –23 and –24. Thus, 21 residues at the C-terminus are exposed to the aqueous environment.

We have determined the amino acid sequence of the N-terminal 83 residues and C-terminal 39 residues containing these three exposed regions (Fig. 1). We note four differences between our results and those published elsewhere¹⁰. These are listed in the legend to Fig. 1. The N-terminal segment is very hydrophobic, and taken in conjunction with the results of the protease digestions of the membrane above, suggests that a section of 63 amino acids—residues 8–70—is buried in the lipid bilayer. The remaining problem is to decide how this buried segment crosses the membrane and whether a structure consistent with the seven-helix model can be proposed. The most

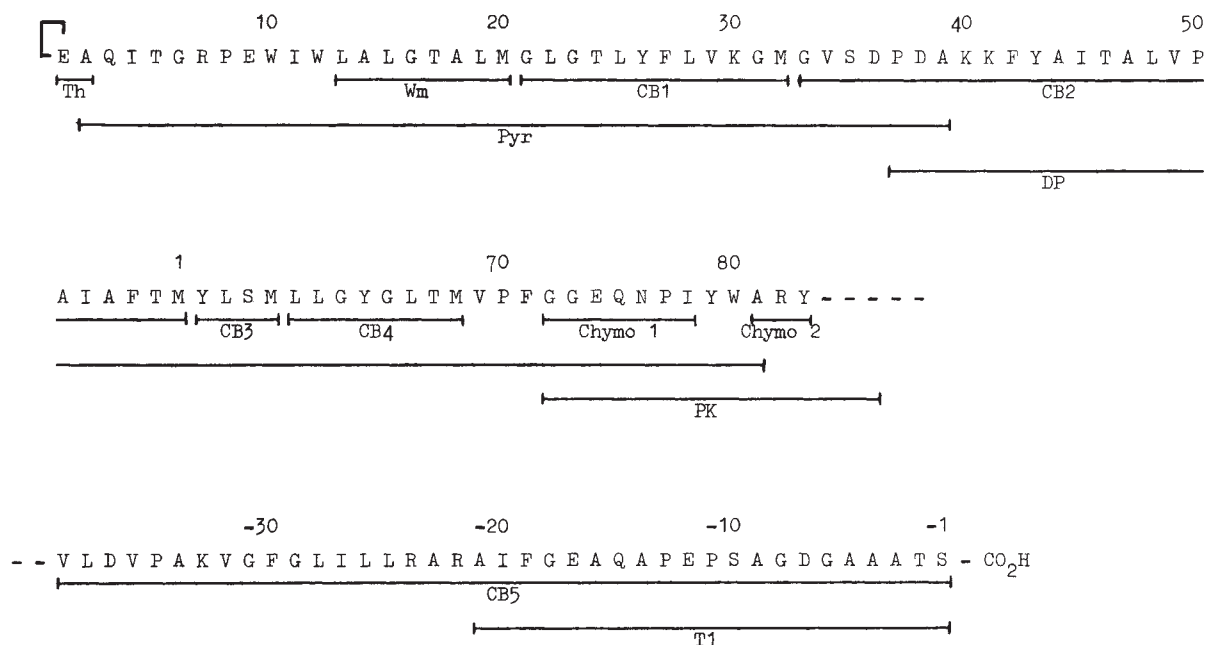


Fig. 1 The amino acid sequence of the N-terminal and C-terminal regions of bacteriorhodopsin. Peptides were derived from digests of the protein with trypsin (T), thermolysin (Th), chymotrypsin (Chymo), cyanogen bromide (CB), cyanogen bromide in 88% formic acid and heptafluorobutyric acid (Wm)¹¹. Enzyme digests were carried out on succinylated protein in 0.5% ammonium bicarbonate, and chemical cleavages on delipidated protein. In addition, the linkage between residues 37 and 38 was cleaved specifically by incubation of delipidated protein in 70% formic acid at 37 °C for 48 h. The pyroglutamyl moiety was removed in 70% yield from succinyl protein by incubation for 30 h at 37 °C with pyroglutamyl amino peptidase (Boehringer) with a substrate : enzyme ratio of 50 : 1. The peptide PK was obtained from a digest of the membrane with proteinase K (4 h, 37 °C). T1 was liberated from the membrane by tryptic digestion. Peptide Th was isolated according to Bruton and Hartley¹². Wm, CB3, CB4, Chymo 1 and Chymo 2 by paper electrophoresis and chromatography, CB1, CB2, CB5, DP and PK by gel filtration of the digests in 70% formic acid on Biogel P-30. The sequences of Pyr, DP, PK, CB1, CB2 and CB5 were obtained with the aid of a spinning cup sequencer¹³ and CB1, CB2, CB3 and CB4 with a solid phase sequencer after coupling to amino-propyl glass¹⁴. Th was sequenced by mass spectrometry, Wm, Chymo 1, Chymo 2 and T1 by the dansyl-Edman procedure¹⁵. Ovchinnikov *et al.*¹⁶ found threonine at position 47, serine at –35, serine at –23 and glutamic acid at –14.

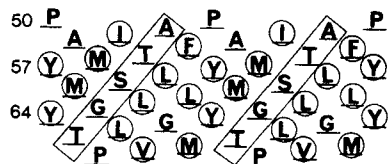


Fig. 2 Amino acid sequence of residues 50–70 plotted on a helix surface with 3.5 residues per turn. Sequence reads downwards right to left as if projected on to a sheet of paper wrapped twice round the helix axis and then unrolled. Each residue appears twice. Hydrophobic residues are circled. The box represents a relatively hydrophilic groove on the helix surface. This is consistent with a helical structure¹⁸.

probable interpretation is that the polypeptide crosses the membrane twice in two helices. Prolines at 8 and 37 are consistent with the notion that residues 9–34 (25 amino acids) form a helix which would span a distance of approximately $25 \times 1.5 \text{ \AA} = 37.5 \text{ \AA}$. This section is terminated by the sequence -S-D-P-D (residues 35–38) which could readily form a β -turn. Residues 39–70 would then cross the membrane a second time, emerging from it at about residue 70. This segment can be considered in two parts: residues 39–49, containing the retinal binding site, and the section between proline residues 50–70. It is likely that this latter section also forms an α -helix. This view is supported by plotting this sequence on a helix surface (Fig. 2). This shows a groove composed of amino acids with less hydrophobic side chains, a feature characteristic of an α -helix¹⁸.

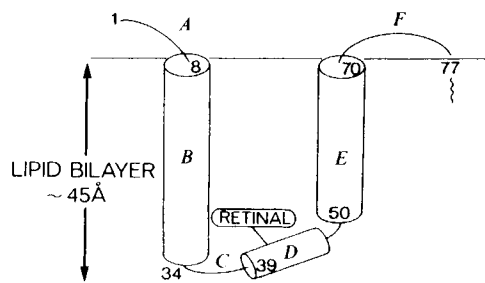


Fig. 3 Proposed topography of N-terminal region of the purple membrane protein. Sections A (residues 1–7) and F (71–76) are accessible to proteolysis. B (residues 8–34) is probably an α -helix, C (residues 34–38) is a β -bend. D (residues 39–49) contains the retinal binding site, E (residues 50–70) may be an α -helix. The angle between the axes of D and E is not considered here. Proline residues occur at positions 8, 37, 50, 70 and 77.

The second helix (residues 50–70, E in Fig. 3) would be about 30 Å long and thus too short to cross the membrane. Hence, residues 39–50 containing the retinal binding site (D in Fig. 3) must partly cross the membrane. The secondary structure of this segment is rather uncertain but it too may be partly α -helical. These features are summarised in Fig. 3. Note that this model places all charged residues near the periphery of the membrane and that none are buried in the interior.

Our model places the N-terminus and loop F on the same side of the membrane. The retinal binding site is near to the opposite side. This is in contrast to earlier models which place loop F on the opposite side from the N-terminus and the retinal lysine variously midway between the two sides of the membrane or near the same side as the N-terminus surface^{9,17}.

Moreover, it has been suggested that the C-terminus is orientated towards the inside of the bacterial cell, and hence in a structure of seven helices it would be outside¹⁰. If this is so the retinal lysine will be near the inside surface of the membrane and loop F will be on the outside.

We thank R. Henderson and A. D. McLachlan for discussions and H. R. Morris for mass spectrometry of N-terminal peptide, Th.

Note added in proof: Recently, Gerber *et al.*¹⁹, using a somewhat different methodology, have determined sequences covering essentially the same regions of the polypeptide chain of bacteriorhodopsin as described here.

JOHN E. WALKER
ALEXANDER F. CARNE
HERBERT W. SCHMITT*

*MRC Laboratory of Molecular Biology,
Hills Road, Cambridge, UK*

Received 17 January; accepted 12 March 1979.

* Present address: Sandoz, 4002 Basel, Switzerland.

- Oesterhelt, D. *FEBS Lett.* **64**, 20–22 (1976).
- Stoeckenius, W. *Scient. Am.* **264**, 38–46 (1976).
- Henderson, R. A. *Rev. Biophys. Bioengng* **6**, 87–109 (1977).
- Bridgen, J. & Walker, J. D. *Biochemistry* **15**, 792–797 (1976).
- Stoeckenius, W. *Biochim. biophys. Acta* (in the press).
- Henderson, R. & Unwin, P. N. T. *Nature* **257**, 28–32 (1975).
- Henderson, R. *J. molec. Biol.* **93**, 123–138 (1975).
- Long, M.M., Urry, D. W. & Stoeckenius, W. *Biochem. biophys. Res. Commun.* **75**, 725–731 (1977).
- Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu, Kiselev, A. V. & Lobanov, N. A. *FEBS Lett.* **84**, 1–4 (1977).
- Gerber, E. G., Gray, P. C., Wildenauer, D. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5426–5430 (1977).
- Ozols, J. & Gerard, C. *J. biol. Chem.* **252**, 5986–5989 (1977).
- Bruton, C. J. & Hartley, B. S. *J. molec. Biol.* **52**, 165–178 (1970).
- Walker, J. E., Carne, A. F., Runswick, M. J., Bridgen, J. & Harris, J. I. *Eur. J. Biochem.* (in the press).
- Wachter, E., Hofner, H. & Machleidt, W. *Solid Phase Methods in Protein Sequence Analysis* (ed. Laursen, R. A.) 31–46 (Pierce Chemicals, Illinois, 1975).
- Hartley, B. S. *Biochem. J.* **119**, 805–822 (1970).
- Ovchinnikov, Yu. A. *et al. Bio-org. Khim.* **4**, 1573–1574 (1978).
- Ovchinnikov, Yu. A. *et al. Bio-org. Khim.* **4**, 979–981 (1978).
- Lim, V. I. *J. molec. Biol.* **88**, 857–894 (1974).
- Gerber, G. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 227–231 (1979).

Globin gene deletion in HPFH, $\delta^{\circ}\beta^{\circ}$ thalassaemia and Hb Lepore disease

THE thalassaemias are a group of inherited disorders characterised by the defective production of either α (α thalassaemias) or non- α (β and $\delta^{\circ}\beta^{\circ}$ thalassaemias) globin chains of haemoglobins (Hb)¹. In β thalassaemias the decreased synthesis of β -globin chains is only partially compensated by the increased production of γ chains, which probably reflects² the massive hypertrophy of the erythron with selective survival of the clones of adult haemoglobin F-producing cells (F cells^{3,4}). The situation is very different in other genetic disorders of the non- α gene cluster, known as $\delta^{\circ}\beta^{\circ}$ thalassaemias and Negro type of hereditary persistence of fetal haemoglobin (HPFH). In these two forms there is a genuine increase of γ -chain production, as shown by the high level of HbF found in heterozygotes. Although a clearcut distinction from both the clinical and haematological point of view cannot be traced between these two forms, the HPFH differs from the $\delta^{\circ}\beta^{\circ}$ thalassaemia in having a higher degree of γ -chain synthesis and a more homogeneous distribution of HbF within red cells. Recently, it has been possible to carry out gene analysis on DNA prepared from β° , $\delta^{\circ}\beta^{\circ}$ thalassaemic and HPFH patients. The β -globin gene is present in β° thalassaemias^{5–9}, but in $\delta^{\circ}\beta^{\circ}$ thalassaemias and HPFH a major deletion, possibly involving both δ and β genes, has been demonstrated by hybridisation studies^{8,10–12}. To characterise the molecular defect in these genetic disorders more precisely, we have hybridised DNA from homozygotes with HPFH, $\delta^{\circ}\beta^{\circ}$ thalassaemia and Hb Lepore disease (in which non- α -chains are a $\delta\beta$ fusion product¹³). For this we used a pure full-size cDNA _{β} probe and specific 5' end and 3' end cDNA fragments (we designate as 5' end cDNA the portion corresponding to the 5' end of the mRNA; the same for the 3' end). Our results, reported here, show that in contrast to HPFH, where a complete δ and β gene deletion occurs, in $\delta^{\circ}\beta^{\circ}$ thalassaemia a 5'-end fragment of the δ gene is present.

Figure 1 shows the titration of β -globin sequences in DNA from a normal human, a case of Hb Lepore, a southern Italian patient with newly discovered δ^β thalassaemia and two patients with HPFH, Negro type (all homozygotes). The cDNA $_\beta$ probe used was a full-size β band eluted from an agarose gel (see legend to Fig. 1), further purified to over 98% (not shown) by hybridisation with mRNA followed by hydroxyapatite (HAP) chromatography⁵. With HPFH DNA no hybridisation could be obtained (except at high cDNA-DNA ratios, when minor contaminating sequences are expected to hybridise^{14,15,22}), in agreement with Orkin *et al.*²³; in contrast, δ^β and Lepore DNA showed a plateau level corresponding to approximately 25–30 and 30–40%, respectively (over the background HPFH DNA), of the hybridisation obtained with normal DNA. These data suggest that part of the $\delta\beta$ sequences are retained in δ^β DNA.

To locate the residual fragment we used 5'-end and 3'-end cDNA probes. A 142-nucleotide long 5'-end fragment (nucleotides 81–222 (ref. 21)) of the β -globin sequence was obtained by digesting the full-size β -globin band with *Hae*III and eluting the heaviest fragment (H1 (ref. 21)) from a preparative polyacrylamide gel. This fragment was further purified from an identical size (141 nucleotides) α sequence²⁴ by hybridisation

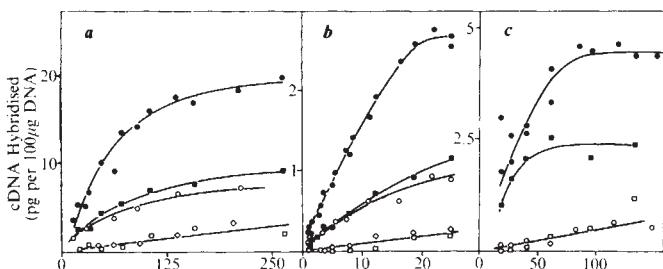


Fig. 1 *a*, Hybridisation of full-size cDNA $_\beta$ to normal, Lepore, δ^β and HPFH DNA in cDNA excess^{5,14,15}. cDNA was copied with avian myeloblastosis virus reverse transcriptase in the presence of actinomycin D from globin mRNA from an HbH disease (α thalassaemia) patient, as described previously¹², except that ³²P-dCTP (Amersham, 400 Ci mmol⁻¹) was 300 μ M, and incubation was carried out at 41 °C for 45 min. The cDNA was then separated from unreacted nucleotides by G50 chromatography, incubated with 0.3 M NaOH at 37 °C for 15 min to break the RNA, neutralised and ethanol precipitated with 40 μ g ml⁻¹ carrier *Escherichia coli* tRNA, pelleted, resuspended in 10 μ l of 1/10 E buffer¹⁶ containing 0.04% bromophenol blue, and electrophoresed at 120 V on a 2% agarose slab gel (15 cm $\times$ 15 cm $\times$ 0.2 cm) in E buffer, until the dye was 1 cm from the bottom of the gel. The gel was then autoradiographed using a Kodak Royal X-0 Mat X-ray film and the heaviest radioactive band (full-size cDNA $_\beta$ (ref. 20)) eluted electrophoretically from slices cut from the gel using the developed film as a reference. All DNAs were prepared by a HAP procedure¹⁷ and fragmented to 200–300 nucleotides by boiling for 20 min in 0.3 M NaOH (ref. 18). Normal and Lepore DNA were from the spleens of an anaemic non-thalassaemic patient and a Lepore Boston disease homozygote, respectively. HPFH DNA 1 was from white blood cells from the recently described homozygote found in Ghana¹⁹ and HPFH DNA 2 was from cultured fibroblasts from the Baltimore homozygote¹². δ^β thalassaemic DNA was from white cells from a newly diagnosed southern Italian patient. DNAs (10–50 μ g) were mixed with increasing amounts of the full-size cDNA $_\beta$ probe in 0.12 M Na phosphate buffer, pH 6.8, 1 mM EDTA, at a DNA concentration of 5 mg ml⁻¹. The aliquots were sealed in silicon-coated glass capillaries, boiled for 10 min and incubated at 70 °C for 4,400 min. At the end of the incubation, the samples were assayed for S₁ nuclease-resistant radioactivity as described elsewhere⁵. The amount of cDNA $_\beta$ hybridised is converted to pg cDNA $_\beta$, and corrected for the background S₁ nuclease resistance (0.6%). Control experiments with cDNA $_{\alpha\gamma}$ show that all these DNAs hybridise identically. Normal, $\bullet$; Lepore, $\blacksquare$; δ^β thalassaemia, $\circ$; HPFH 1, $\square$; HPFH 2, $\diamond$. Hybridisation of 5' cDNA $_\beta$ fragment (*b*) and 3' cDNA $_\beta$ fragment (*c*) to normal, Lepore, δ^β , HPFH 1 and HPFH 2 DNA. Full size cDNA $_\beta$ (ref. 20) eluted from the agarose gel (see above) was ethanol precipitated using 20 μ g ml⁻¹ Dextran (Sigma) as a carrier. After pelleting, the cDNA was resuspended and digested with 100 units μ g⁻¹ of *Hae*III restriction endonuclease in 10 mM Tris HCl, pH 7.6, 10 mM MgCl₂, 1 mM dithiothreitol for 4 h at 37 °C. The digested sample was directly loaded on a 7.5% polyacrylamide gel in E buffer and the separated bands located and recovered as described above. All the probes used for the hybridisation were further purified from contaminating α sequences by annealing to excess $\alpha\gamma$ mRNA and HAP chromatography⁵. DNAs were mixed with increasing amounts of *Hae*III H 1 (ref. 21) cDNA $_\beta$ fragment (*b*), or 3' end cDNA $_\beta$ probe (*c*), and hybridised as above for 5,000 min. A background S₁ nuclease resistance of 0.4% was subtracted.

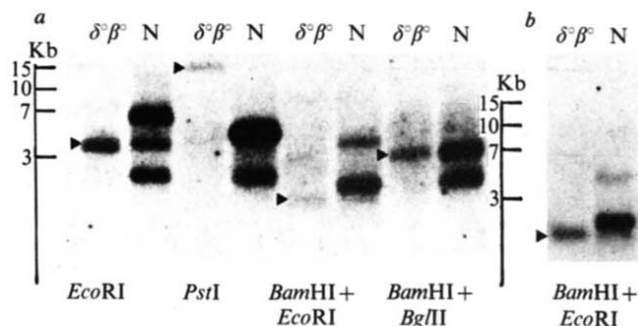
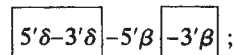


Fig. 2 Cleavage pattern of δ and β globin genes with different restriction enzymes. High molecular weight DNA was prepared according to Britten *et al.*³⁰ from normal human placenta or from purified white cell nuclei³¹ of the δ^β homozygous patient. DNA (30 μ g) was digested with the indicated restriction enzymes, denatured with alkali^{32,33} and electrophoresed in 0.8% agarose gel in E buffer¹⁶. DNA transfer on to nitrocellulose filters and hybridisation at 65 °C in 2 $\times$ SSC was according to Jeffreys and Flavell^{32,33} using nick-translated³² pH β G1 plasmid as a radioactive probe (specific activity 50–300 $\times$ 10⁶ c.p.m. μ g⁻¹). Filter washing was carried out in 0.1 $\times$ SSC (*a*) at 65 °C for 6 h with vigorous shaking. This washing procedure seems to be more stringent than that used by Flavell *et al.*^{29,32,33} at the same salt concentration, as shown by the faintness of δ bands in 0.1 $\times$ SSC (*a*) and the absence of γ bands even at 0.3 $\times$ SSC (*b*).

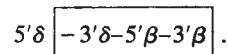
with mRNA $_{\alpha\gamma}$ and HAP fractionation⁵. Figure 1*b* shows that this 5'-end fragment does not hybridise to HPFH DNA (as expected), whereas it does hybridise to a comparable extent to both Lepore and δ^β DNA (30% and 40%, respectively, of the normal). As the non- α chain of this case of Hb Lepore is a Boston type (in which a change-over from the δ -like sequence to the β -like sequence occurs between amino acids 87 and 116 (refs 25, 26)), this result demonstrates that this portion of the cDNA $_\beta$ can cross-hybridise to the corresponding δ sequence.

A 3'-end probe was obtained by eluting from an agarose gel a partial transcript (of approximately 200–250 nucleotides) obtained by reverse transcription of α -thalassaemic mRNA. In agreement with Efstradiatis (see ref. 27, and results not shown), we find that these fragments are incomplete oligo(dT)-containing products initiated at the 3'-poly(A) tail of the template and terminated at defined internal points. In contrast to the results obtained with the 5'-end probe, this fragment (further purified as above by HAP chromatography) does not hybridise with δ^β DNA; the same was obtained with HPFH DNA, whereas the Lepore DNA hybridised to approximately 40–50% of normal (Fig. 2*c*), in agreement with the expected absence of the 3' end of the δ gene (which in these conditions seems to cross-hybridise efficiently to the probe used).

Two different schemes can fit our data; first, that the 5' end of the β -globin gene is retained in δ^β thalassaemia, whereas the whole of the δ gene and the 3' end of the β -globin gene are missing:



alternatively, only the 5' δ -globin fragment would be present:



To distinguish between these two possibilities, we carried out mapping experiments using as a probe nick-translated radioactive pH β G1 plasmid containing human β -globin sequences²⁸.

As shown in Fig. 2, on digestion with *Eco*RI, δ^β DNA shows, instead of the three main bands of 6.5, 4 and 2.3 kilobases present in normal DNA²⁹, a single band running slightly faster than the 4-kilobase β fragment; this result agrees with that reported by Mears *et al.* in a different patient³⁴. Double digestion with *Bam*HI and *Bgl*II results in a single 3.3-kilobase fragment in the δ^β DNA; in normal DNA this fragment can originate both from the 3' end β gene and from the 5'-end δ gene. Thus, two alternative schemes can be proposed: the first

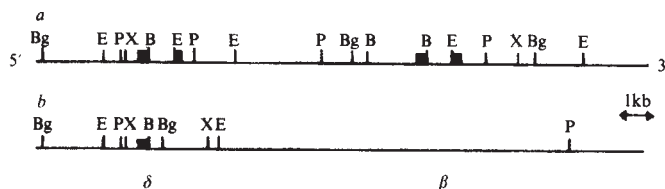


Fig. 3 Physical maps of δ and β DNA region in normal (a) and thalassaemic (b) DNA. Normal map is from Flavell *et al.*²⁹; $\delta^{\circ}\beta^{\circ}$ map was deduced from the digests reported in Table 1. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; P, *Pst*I; X, *Xba*I. The position of the *Pst*I site 5' to the δ gene in $\delta^{\circ}\beta^{\circ}$ patient is based mainly on its location in normal DNA and indirectly deduced from the persistence of the flanking *Eco*RI and *Xba*I sites.

implies the deletion of the β -globin gene and possibly part of the δ gene; the second requires deletion of the δ gene and the 5' end of the β -globin gene. The demonstration of the normal 1.3-kilobase *Bam*HI + *Eco*RI fragment (Fig. 2 and Table 1) and the 0.7-kilobase *Bam*HI + *Xba*I fragment (Table 1) in the $\delta^{\circ}\beta^{\circ}$ DNA confirms the validity of the first scheme. The experiment shown in Fig. 2 was carried out in highly stringent conditions (see Fig. 2 legend), a fact that explains the faintness of the δ bands. As expected, the use of less stringent washing conditions (Fig. 2b) results in a striking increase in the intensity of these bands (nevertheless, these conditions are stringent enough to prevent the cross-hybridisation of γ genes to the β probe, as confirmed by control experiments with HPFH DNA, not shown).

From the liquid hybridisation and restriction mapping experiments, we conclude that: (1) a 5'-end fragment is retained in $\delta^{\circ}\beta^{\circ}$ thalassaemia (Fig. 1b) which can be identified as the 5' end of the δ gene (Fig. 2 and Table 1); (2) no δ or β 3'-end fragments are retained in $\delta^{\circ}\beta^{\circ}$ DNA, as shown by liquid hybridisation methods (Fig. 1c), by the absence of the normal 4-kilobase fragment in the *Eco*RI + *Bam*HI double digest, and by the disappearance of the 2.3-kilobase δ fragment expected if the *Eco*RI site present in the δ gene-coding region were retained; (3) from the data in Table 1 we have constructed the physical map (Fig. 3) of the $\delta\beta$ -globin region in this $\delta^{\circ}\beta^{\circ}$ thalassaemic DNA. It is clear that the deletion removes a large DNA fragment beyond the δ intragenic *Bam*HI site, extending past the first *Eco*RI site, on the 3' side of the β -globin gene region: in fact, the relative positions of the *Pst*I, *Xba*I, *Bgl*II and *Eco*RI sites in this region do not correspond to the normal ones.

Genetic data based on the analysis of the non- α chain of Hb Kenya (a $\gamma\beta$ fusion product^{35,36}) and Hb Lepore ($\delta\beta$ fusion product) indicate that the most likely arrangement of the non- α genes on the chromosome is $\gamma^{\circ}\gamma^{\circ}\delta^{\circ}\beta^{\circ}$. Because Hb Kenya mutation is associated with high levels of HbF in adult life, whereas usually the Hb Lepore is not, a 'deletion hypothesis' (refs 37, 38) has been proposed to explain the molecular basis of HPFH. According to this model, the deletion in HPFH involves an area of the $\gamma\delta\beta$ -gene complex (probably located to the left of the δ gene; Y area of Huisman *et al.*³⁷) which is crucial to the post-natal suppression of γ -chain production; removal of this DNA segment allows the combined expression of γ genes, although not to the extent that they function in fetal life.

Our results only partially agree with this model, as a significant de-repression of γ genes occurs even in the presence of an intact 5'-end region of the δ gene. Thus, the structural δ -gene sequences and/or the insert might be part of the postu-

lated regulatory area controlling γ -gene expression. On the other hand, a larger deletion, involving an additional region including the 5' end of the δ gene, results in the HPFH picture. Establishment of the exact DNA region involved in determining the difference between the HPFH and $\delta^{\circ}\beta^{\circ}$ thalassaemia picture, will require mapping studies of non-structural gene areas at the 5' end of the δ gene in additional patients (see Orkin *et al.*²³). So far, our data are consistent with the hypothesis of Huisman *et al.*³⁷ that a continuum of molecular lesions might exist between two extremes: the HPFH syndrome with maximum de-repression of γ -chain synthesis and deletions extending from the β -globin gene to the γ gene; and the $\delta^{\circ}\beta^{\circ}$ thalassaemia with minor de-repression of the γ -globin genes and a more limited gene deletion.

We thank Drs Peter Little and Bob Williamson for the gift of pH β G1 and pH γ G1 plasmids, Professor L. Luzzatto for stimulating the collaboration between Kumasi and Milano Universities, Drs N. Cattaneo and A. Rossi for providing HbH blood samples, and Drs G. Lanyon and I. Pragnell for sending HPFH DNA. This work was partially supported by the CNR grant CT 7701451.04 to E.P., and a grant from Fondazione Anna Villa Rusconi, Varese.

S. OTTOLENGHI
B. GIGLIONI
P. COMI

Istituto di Patologia Generale,
Centro per lo Studio della Patologia
Cellulare del CNR,
Università di Milano

A. M. GIANNI
E. POLLI

Istituto di Clinica Medica I (Granelli),
Università di Milano

C. T. A. ACQUAYE
J. H. OLDHAM

Department of Biochemistry,
Kumasi University, Ghana

G. MASERA

Istituto di Clinica Pediatrica,
Università di Milano, Italia

Received 26 September 1978; accepted 19 February 1979.

- Weatherall, D. J. & Clegg, J. B. *The Thalassemia Syndromes* 2nd edn (Blackwell, Oxford, 1972).
- Weatherall, D. J. *Congenital Disorders of Erythropoiesis* (Ciba Symp., North-Holland, Amsterdam, 1976).
- Boyer, S. H., Belding, T. K., Margolet, L. & Noyes, A. N. *Science* **188**, 361-363 (1975).
- Wood, W. G., Stamatoyannopoulos, G., Lim, G. & Nute, P. E. *Blood* **46**, 671-682 (1975).
- Ottolenghi, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 2294-2299 (1975).
- Kan, Y. W., Holland, J. P., Dozy, A. M. & Varmus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5140-5144 (1975).
- Tolstoshev, P. *et al. Nature* **259**, 95-98 (1976).
- Ramirez, F. *et al. Nature* **263**, 471-475 (1976).
- Comi, P. *et al. Eur. J. Biochem.* **79**, 617-622 (1977).
- Kan, Y. W., Holland, J. P., Dozy, A. M., Charache, S. & Kazazian, H. H. *Nature* **258**, 162-163 (1975).
- Forget, B. G. *et al. Cell* **7**, 323-329 (1976).
- Ottolenghi, S. *et al. Cell* **9**, 71-80 (1976).
- Baglioni, C. *Proc. natn. Acad. Sci. U.S.A.* **48**, 1880-1886 (1962).
- Bishop, J. O. & Freeman, K. B. *Cold Spring Harb. Symp. quant. Biol.* **38**, 707-716 (1974).
- Williamson, R. *Br. med. Bull.* **32**, 246-252 (1976).
- Sharp, P. A., Sugden, B. & Sambrook, J. *Biochemistry* **12**, 3055-3063 (1973).
- Hell, A., Birnie, G. D., Slimming, T. K. & Paul, J. *Analyt. Biochem.* **48**, 369-377 (1972).
- Tibbets, C., Johansson, K. & Philipson, L. *J. Virol.* **12**, 218-225 (1973).
- Acquaye, C. T. A., Oldham, J. H. & Konotey-Ahulu, F. I. D. *Lancet* **i**, 796-797 (1977).
- Orkin, S. H. *J. Biol. Chem.* **253**, 12-15 (1978).
- Marotta, C. A., Wilson, J. T., Forget, B. G. & Weissman, S. M. *J. Biol. Chem.* **252**, 5040-5053 (1977).
- Old, J. *et al. Cell* **8**, 13-18 (1976).
- Orkin, S. *et al. New Engl. J. Med.* **299**, 166-174 (1978).
- Wilson, J. T., DeRiel, J. K., Forget, B. G., Marotta, C. A. & Weissman, S. M. *Nucleic Acids Res.* **4**, 2353-2368 (1977).
- Baglioni, C. *Biochim. biophys. Acta* **97**, 37-46 (1965).
- Labie, D., Schroeder, W. A. & Huisman, T. H. J. *Biochim. biophys. Acta* **127**, 428-437 (1966).
- Elstradiatis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A. & Vournakis, J. N. *Cell* **4**, 367-378 (1975).
- Little, P. *et al. Nature* **273**, 640-643 (1978).
- Flavell, R. A., Kooter, J. M., DeBoer, E., Little, P. F. R. & Williamson, R. *Cell* **15**, 25-38 (1978).
- Britten, R. J., Graham, D. E. & Neufeld, B. R. *Meth. Enzym.* **29**, part E, 363-418 (1974).
- Gianni, A. M. *et al. Nature* **274**, 610-612 (1978).
- Jeffrey, A. J. & Flavell, R. A. *Cell* **12**, 429-439 (1977).

Table 1 Size of β -like globin fragments in $\delta^{\circ}\beta^{\circ}$ human thalassaemic DNA cleaved by various restriction enzymes

<i>Pst</i> I	<i>Bgl</i> II	<i>Eco</i> RI	<i>Xba</i> I	<i>Bam</i> HI + <i>Xba</i> I	<i>Bam</i> HI + <i>Eco</i> RI	<i>Eco</i> RI + <i>Xba</i> I	<i>Bam</i> HI + <i>Bgl</i> II
15	4	3.9	2.8	0.7	1.3	2.8	3.3

Size (in kilobases) was estimated by comparison with β -like DNA fragments generated by cleavage of normal DNA with *Eco*RI (6.5, 4, 1.3 kilobases), *Pst*I (5, 2.3 kilobases) *Bam*HI + *Bgl*II (3.3, 1.8 kilobases) and *Bam*HI + *Eco*RI (4, 1.8, 1.3 kilobases) and run in parallel lanes (for size, see ref. 29).

33. Jeffreys, A. J. & Flavell, R. A. *Cell* **12**, 1097–1108 (1977).
 34. Mears, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 1222–1226 (1978).
 35. Huisman, T. H. J., Wrightstone, R. N., Wilson, J. B., Schroeder, W. A. & Kendall, A. G. *Archs Biochem. Biophys.* **153**, 850–853 (1972).
 36. Smith, D. H., Clegg, J. B., Weatherall, D. J. & Gilles, H. M. *Nature new Biol.* **246**, 184–186 (1973).
 37. Huisman, T. H. J. *et al. Ann. N.Y. Acad. Sci.* **232**, 107–124 (1974).
 38. Wood, W. G., Clegg, J. B. & Weatherall, D. J. *Prog. Haemat.* **43**–90 (1977).

Common regulatory mechanism of expression and conjugative ability of a tetracycline resistance plasmid in *Bacteroides fragilis*

RESISTANCE to various antibiotics is common in the anaerobic gut bacterium *Bacteroides fragilis*. A transferable plasmid coding for resistance to antibiotics of the lincosamide/macrolide/synergistin groups has been found^{1,2} and transfer is most probably a conjugative process. This supports the observation of the recent emergence of resistance to antibiotics in these groups and its rapid spread in certain areas (ref. 3 and our unpublished data). However, in the case of tetracycline resistance the limited genetic and physical data were at variance with the epidemiological data. Although tetracycline resistance is common in *B. fragilis*⁴ (at present 60% of strains isolated in the USA⁵ and 30% of those isolated in France (our unpublished results) are resistant) intraspecies transfer of tetracycline resistance could not be demonstrated⁶. We report here that transfer of tetracycline resistance between *B. fragilis* strains is easily achieved if the donor bacteria are grown in the presence of tetracycline before mating. This suggests that tetracycline resistance is borne on a transferable plasmid whose conjugative transfer requires activation of the tetracycline resistance gene. This is supported by our finding that the transfer of tetracycline resistance is either inducible or constitutive, depending on whether tetracycline resistance itself is respectively inducible by tetracycline or constitutive.

At first we failed to transfer Tc resistance from *B. fragilis* strain 92 to a sensitive plasmid-free *B. fragilis* strain (638 rfm); using the same recipient we occasionally succeeded in transferring Tc resistance from a *B. distasonis* donor strain (BEN) at very low frequency¹. Nevertheless, with both donor strains, erythromycin-clindamycin (Em-Cl) resistance was easily shown to be transferable. The plasmid character of resistance to the

macrolides, the lincosamines, and to the synergistins was obvious because they were curable *en bloc* and were easy to characterise in strain 92 and transcipts as DNA molecules of molecular weight between 27,000 (ref. 2) and 28,000 (our unpublished results). In similar experiments we failed to eliminate the Tc resistance¹ but the extrachromosomal localisation of Tc resistance was shown in an Em-Cl cured derivative of strain 92 and in Tc transcipts (plasmid of 17,000 molecular weight).

It seemed possible that the rapid spread of Tc resistance in *B. fragilis* could be due to a plasmid whose transfer requires certain conditions present in its usual human or animal habitat (the intestine), but not in the normal experimental situation. Among the possible factors 'necessary' for this transfer, a good candidate was tetracycline itself which is widely used in therapy and in animal foods.

In strain 92, tetracycline resistance is inducible by tetracycline if the bacteria are exposed to the antibiotic for 1 h or more (Fig. 1); this strain is said to be Tcⁱ. The mechanism of plasmid-mediated Tc resistance is not known in *B. fragilis*. But we can reasonably assume that it is under negative regulatory control, as is inducible Tc resistance in facultative anaerobic bacteria⁷.

When strain 92 was cultivated for several generations in the presence of subinhibitory concentrations of tetracycline and then immediately mated with the recipient, Tc transcipts were obtained. Results of a typical experiment are given in Table 1, and similar results were obtained in other serial subculture experiments. The number of Tc transcipts recovered increased during these initial subcultures to 10⁵–10⁷ ml⁻¹ (Table 1, column 3), and at the same time the number of Em-Cl transcipts was relatively stable, or decreased by spontaneous curing of the Em-Cl plasmid. The ratios of Tc^R transcipts to Em^R transcipts are indicated in Table 1 as an approximate measure of Tc transfer efficiency.

When the donor strain was in contact with tetracycline for 15 generations or less, its ability to transfer Tc resistance was strictly inducible (Tra_i), as it disappeared when the Tc-induced cultures were serially subcultured in the absence of tetracycline for ≥120 generations (de-induction). The Tc resistance of these de-induced cultures remained inducible (Tc_i^R), as in the original strain. In contrast to these results, data from Table 1 show that after more generations (≥20) of the donor strain in the presence of tetracycline, its ability to transfer Tc resistance was maintained after de-induction; the transfer ability appeared to be constitutive (Tra_c). Concomitantly, the Tc resistance itself turned out to be constitutive (Tc_c^R). After a greater number of

Table 1 Inducibility against constitutivity of Tc transferability and Tc resistance in *B. fragilis* strain 92

Donor strain*	Tetracycline concentration used before mating (μg ml ⁻¹)	Tc	Induced conjugative ability (No. transcipts per ml of suspension)†		Tc/Em	Inducible versus constitutive Tc ^R transfer‡	Inducibility versus constitutivity of Tc resistance§
			Em				
92	0	0	3 × 10 ⁵		<3 × 10 ⁻⁴	—	i
92-Tc ₅	10	2 × 10 ⁵	2 × 10 ⁵		1	i	i
92-Tc ₁₀	10	3 × 10 ⁵	1 × 10 ⁵		3	i	i
92-Tc ₁₅	40	4.5 × 10 ⁷	5 × 10 ³		~10 ⁴	i	i
92-Tc ₂₀	40	+	+		—	i and c	i and c
92-Tc ₂₅	40	+	+		—	c	c

* The donor strain was inoculated at 4 × 10⁷ bacteria ml⁻¹ in TYG broth¹ in the presence of variable concentrations of tetracycline. Subcultures were similarly made daily from the higher non-inhibitory Tc concentration. In these conditions each subculture is equivalent to five generations. After 5, 10, ... generations in the presence of Tc, strain 92 is designated as 92-Tc₅, 92-Tc₁₀ and so on.

† Conjugative ability was tested after each subculture in the presence of Tc, by mating the donor with *B. fragilis* 638 rfm according to ref. 1; 0.05 ml of each strain were mixed onto TYG plates of 5 cm diameter. After an overnight incubation the culture was scraped, suspended in 2 ml of TYG and various dilutions of the suspension spread onto agar plates supplemented with rifampicin (Rfm) at 25 μg ml⁻¹ and either tetracycline (Tc) at 2 μg ml⁻¹ or erythromycin (Em) at 10 μg ml⁻¹. The numbers of transcipts per ml of this suspension are indicated in column 3. Isolation and characterisation of transcipts, and conditions of growth were as in ref. 1.

‡ After each subculture in the presence of Tc, the culture was de-induced by growth without Tc for at least 120 generations and tested for its transferability as before.

§ After de-induction, the cultures were tested for inducibility (i) versus constitutivity (c) of Tc resistance using the conditions in Fig. 1, except that the induction time was routinely 18.

serial subcultures (between 40 and 100 generations) in the presence of tetracycline, identical results were obtained, but the minimal inhibitory concentrations (MIC) of Tc for the donor were significantly higher ($100 \mu\text{g ml}^{-1}$) than initially ($16 \mu\text{g ml}^{-1}$), in contrast to identical values during the initial serial subcultures (MIC Tc, $20\text{--}40 \mu\text{g ml}^{-1}$).

In the experiment described in Table 1, the de-induced donor 92-Tc₂₀ which was at first phenotypically Tra_c turned out to be a mixture of Tc_i^R and Tc_c^R clones; when both clones were rechecked for ability to transfer Tc resistance, they were found to be Tra_i and Tra_c, respectively.

The number of generations in the presence of tetracycline required to obtain constitutive Tc resistance and Tc transfer varied from one serial subculture experiment to another, but the phenomenon was experimentally reproducible between the 10th and 25th generations.

The fact that Tc transfer appeared even after a few generations (5–25) in the presence of tetracycline without significant modification of the MIC of the donor seems to suggest a modification of the conjugative ability itself, rather than quicker or more efficient Tc^R expression in transipients. Two findings support this view. (1) In the mating of strain 92 (without any previous contact with Tc) and of the recipient strain 638 rf^m, no Tc transipients were obtained even when the Tc concentration was lowered in the selective plates to $0.2 \mu\text{g ml}^{-1}$ instead of $2 \mu\text{g ml}^{-1}$ as used in all experiments previously reported (MIC Tc of the recipient strain, $0.064 \mu\text{g ml}^{-1}$). (2) In a similar mating using as donor a Tra_c Tc_c^R strain obtained by 20 serial subcultures at a sub-inhibitory concentration of tetracycline (MIC Tc, $100 \mu\text{g ml}^{-1}$) the same number of transipients was obtained by using for the selection different concentrations of tetracycline (Table 2).

We verified that the Tc plasmid transfer was specifically induced by tetracycline or minocycline, but not by erythromycin. The Em-Cl plasmid is not required for Tc plasmid transfer, because identical results were obtained with strain 92-1 which is a derivative, cured of its Em-Cl plasmid¹.

Physical analysis of Tc transipients and of a Tra_c Tc_c^R derivative from strain 92 shows that these strains harbour a plasmid of molecular weight 17,000 as does the wild type strain; consequently plasmid amplification can be ruled out as a possible mechanism for the newly acquired properties.

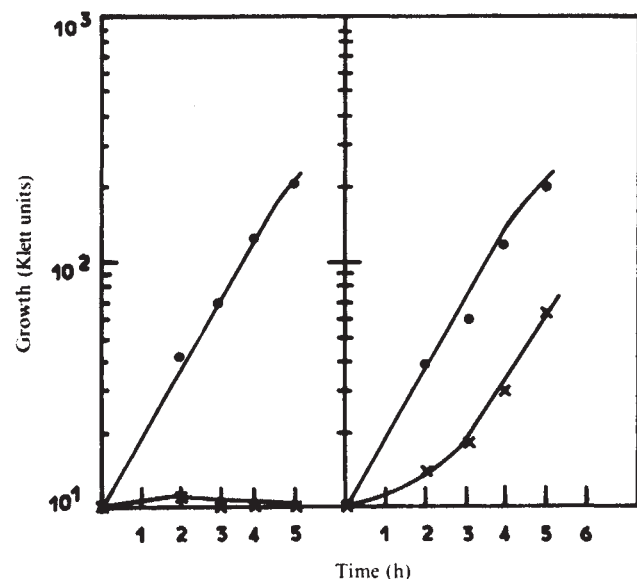


Fig. 1 Induction of tetracycline resistance in *B. fragilis* strain 92. Cultures inoculated with early stationary cells ($10^8 \text{ cells ml}^{-1}$) were grown in TYG broth¹ for 1 h without tetracycline (Tc) (a) or with $0.5 \mu\text{g ml}^{-1}$ of Tc (b). At this time (T_0) they were inoculated at $5 \times 10^6 \text{ cells ml}^{-1}$ in TYG broth without Tc (○) or with Tc at $5 \mu\text{g ml}^{-1}$ (×). Cultures were incubated at 34°C in anaerobic jars and the optical densities were measured at hourly intervals (Klett photocolormeter, filter no. 54).

Table 2 Mating of strain 92-Tc₁₀₀ and strain 638 rf^m*

No. of transipients‡	Tc concentration in the selective plates ($\mu\text{g ml}^{-1}$)†		
	2	20	40
	3.2×10^3	3.0×10^3	2.1×10^3

* The donor strain had a MIC for Tc of $100 \mu\text{g ml}^{-1}$. Before mating it was de-induced by 150 generations in TYG broth in the absence of tetracycline. Other experimental conditions were as in Table 1.

† Selection was made onto TYG agar medium supplemented with rifampicin at $25 \mu\text{g ml}^{-1}$ and tetracycline at the indicated concentration.

‡ The number of transipients is expressed per ml of mating suspension (see Table 1).

Similar results were obtained with other clinical isolates of *B. fragilis*; in these strains, Tc resistance and transfer ability were shown to be inducible by tetracycline in the same experimental conditions. Similarly, when strain BEN of *B. distasonis*¹, also Tc_i^R, was used as a donor, a considerable increase ($>10^3$) in Tc^R transipients was regularly obtained when the strain was subcultured in the presence of tetracycline before mating. The phenotype Tra_i can be attributed to this strain in spite of its very low Tc^R conjugative ability without Tc induction¹. In strain BEN it is probable that an Em-Cl plasmid is co-transferred by Tc plasmid¹. The induction of transferability of Tc plasmid led to a parallel increase of co-transfer. Spontaneous Tra_c Tc_c^R mutants were also obtained from strain BEN.

These results show that growth of a few generations of the Tc_i^R donor strain in the presence of tetracycline reversibly and specifically induces Tc conjugative ability. This suggests induction (or de-repression) of the conjugative ability of the Tc plasmid by tetracycline. At the same time tetracycline is likely to act as the inducer or de-repressor of a product required for expression of Tc resistance. This result is comparable to two reports from studies of metabolic plasmids: Nakasawa and Yokota⁸ showed stimulation of TOL plasmid transfer by *m*-toluate in *Pseudomonas putida*, and Petit *et al.*⁹ demonstrated a stimulatory effect of nopaline or octopine on Ti plasmid transfer in *Agrobacterium tumefaciens*. Our findings with an antibiotic resistance plasmid confirm the prediction of Petit *et al.*⁹ that RTF factors will exist with a conjugative ability inducible by the antibiotic whose resistance they are coding. The induction of plasmid transfer by induction of a gene carried on it appears therefore to be a widespread phenomenon in bacteria. Moreover, with *B. fragilis* strain 92, a greater number of generations grown in the presence of Tc led to Tc_c^R derivatives stably constitutive for transfer (Tra_c) (data not shown). These derivatives are probably spontaneous mutants selected by the antibiotic, whose mutation confers constitutivity for both Tc resistance and Tc^R conjugative ability. The mutation could be in a regulatory gene controlling both properties.

These results lead us to postulate for the Tc plasmids of *B. fragilis* a common regulatory mechanism for resistance and for conjugative ability. These results may help to explain the emergence of Tc^R strains in the intestinal flora during long-term tetracycline therapy. They emphasise the extreme danger of using tetracycline in animal foods or in therapy, unless strictly indicated.

G.P. is supported by a fellowship from the European Molecular Biology Organization.

GAETANO PRIVITERA
MADELEINE SEBALD
FRANÇOISE FAYOLLE

Service des Anaérobies, Institut Pasteur,
75 724 Paris, France

Received 15 December 1978; accepted 19 February 1979.

- Privitera, G., Dublanche, A. & Sebal, M. *J. inf. Dis.* **139**, 97–101 (1979).
- Welch, R. A., Jones, K. R. & Macrina, F. L. *Plasmid*, (in the press).
- Rozmiej, E. A., Bawdon, R. E., Palchaudhuri, S. & Krakowiak, J. *18th. Intersci. Conf. Antimicrob. Ag. and Chemother.*, Atlanta, abstr. 86 (1978).

4. Sutter, V. L., Kwok, Y. Y. & Finegold, S. M. in *Anaerobic Bacteria: Role in Disease* (eds Balows, A., Dehaan, R. M., Dowell, V. R. & Guze, L. B.) (Charles Thomas, Springfield, 1972).
5. Finegold, S. M. in *Anaerobic Bacteria in Human Disease* (ed. Finegold, S. M.) (Academic, New York, 1977).
6. Guiney, D. G. & Davis, C. E. *J. Bact.* **124**, 503–510 (1975).
7. Yang, H. L., Zubay, G. & Levy, S. B. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1509–1512 (1976).
8. Nakazawa, T. & Yokota, T. *J. Bact.* **129**, 39–46 (1977).
9. Petit, A. *et al. Nature* **271**, 570–572 (1978).

Cytochrome c_2 sequence variation among the recognised species of purple nonsulphur photosynthetic bacteria

IN higher organisms, the nature and amount of amino acid sequence difference between homologous proteins from different species correlates well with what is known about the phylogeny of the organisms from morphological and palaeontological evidence^{1–3}. We are interested in the application of amino acid sequence information to the study of bacterial phylogeny, a subject that is not amenable to investigation by the classical methods because of the paucity of morphological features and the lack of a palaeontological record. We have, therefore, determined the amino acid sequences of the principal soluble cytochromes c of representatives of each recognised species of the Rhodospirillaceae, the nonsulphur purple photosynthetic bacteria. In contrast to higher organisms, there was no correlation between a species' cytochrome c amino acid sequence and its phylogenetic position.

The Rhodospirillaceae are the best studied group of phototrophic bacteria^{4,5}, and they have been divided into three genera containing a total of 12 species⁶ (Table 1). The properties of most of these species were found by the study of numerous separate isolates in pure culture. The different species (Table 1) vary widely in their physiology and morphology, and most seem to be definite taxonomic entities. Selected strains of each species have been grown in large quantities and contain high concentrations of a variety of photosynthetic pigments and electron transport proteins⁷, of which the relevant ones are shown in Table 1.

Rhodospirillum rubrum is the type species of the family which produces the well known cytochrome c_2 . Both the amino acid sequence⁸ and the three dimensional structure⁹ are known, and considerable similarity to mitochondrial cytochrome c has been recognised in both sequence and folding. Cytochromes c_2 function in photosynthetic electron transport^{10,11}. They may also function in some modes of non-photosynthetic metabolism, but this is not proved.

We have examined the major, soluble, low-spin c -type cytochromes from 11 other species of Rhodospirillaceae. Most of them produce a protein (or proteins) that resemble cytochrome c_2 and mitochondrial cytochromes c in sequence (Fig. 1). However, two of the species, *Rhodospirillum tenue* and *Rhodopseudomonas gelatinosa*, do not apparently produce any protein of this class but instead form a cytochrome¹² that resembles the small cytochrome c -551 of the denitrifying pseudomonads. The cytochromes c_2 from five of the species (Fig. 1*c–i*) have amino acid sequences resembling those of mitochondrial cytochromes c at least as closely as they do those of *R. rubrum* cytochrome c_2 . These sequences can be aligned with those of the mitochondrial proteins with good matching throughout¹³ (Fig. 1). The remaining four species (Fig. 1*j–o*) produce proteins that resemble *R. rubrum* cytochrome c_2 in that they have appreciably longer sequences than the others (116–124 residues), and to get good matching with the smaller cytochromes c_2 or with mitochondrial cytochromes c require at least three- and eight-residue insertions plus a single-residue deletion. Each of these four species produces sequences containing distinctive insertions and deletions. The cytochrome c -550 from the denitrifying bacterium *Paracoccus denitrificans*^{14,15} also belongs to this sequence group, and shows particular affinity to the *Rhodopseudomonas capsulata* sequence.

We have examined the proteins from two independent strains of *Rps capsulata* and *Rhodopseudomonas palustris*. The former proteins differ from each other in only two positions, while the latter differ in 12 positions. These differences are much smaller than those between any two different species except *Rhodospirillum molischianum* and *Rhodospirillum fulvum*.

*R. molischianum*¹⁶ and *R. fulvum* both produce two distinct soluble cytochromes c_2 . Despite earlier reports¹⁶, both isocytochromes are very similar in size. The sequences of both the *iso*-1-cytochrome c_2 and the *iso*-2-cytochrome c_2 of *R. fulvum* differ from the corresponding proteins of *R. molischianum* in 12 positions. *Rhodospirillum molischianum* and *R. fulvum* are taxonomically similar⁶, the main distinctions between them being cell width *p*-aminobenzoate requirement and ability to utilise benzoate.

The *iso*-1-cytochrome c_2 of both *R. molischianum* and *R. fulvum* lack the 'evolutionarily invariant' tryptophan residue at position 59 (Fig. 1), common to all natural mitochondrial cytochromes c so far examined and to the other cytochromes c_2 shown in Fig. 1. An unstable mutant form of yeast cytochrome c has been produced in which this tryptophan residue has been replaced¹⁷.

The *iso*-2-cytochromes c_2 of both *R. molischianum* and *R. fulvum* seem to contain a mixture of phenylalanine and tyrosine

Table 1 Properties of members of the family Rhodospirillaceae⁶ including occurrence of electron transport proteins⁷

Species	Strain	ATCC no.	Electron transport proteins					Mode of cell division	Type of photosynthetic membrane	Cell form
			Large c_2	Small c_2	c -551	c'	HiPIP			
<i>Rhodospirillum rubrum</i>	S1 (=1.1.1)	11170	+			+		Binary fission	Vesicular	Spiral
<i>R. tenue</i>	3761				+	+	+	Binary fission	Tubular	Thin spiral
<i>R. fulvum</i>	1360	15798		+		+		Binary fission	Lamellar	Short spiral
<i>R. molischianum</i>	S	14031		+		+		Binary fission	Lamellar	Spiral
<i>R. photometricum</i>	SP113		+			+		Binary fission	Lamellar	Large spiral
<i>Rhodopseudomonas</i>	2.1.6	17001	+					Bud	Lamellar	Rod
<i>palustris</i>	2.1.37	17007	+			+		Bud	Lamellar	Rod
<i>Rps viridis</i>	NTHC 133			+				Bud	Lamellar	Rod
<i>Rps acidophila</i>	7050	25092		+				Bud	Lamellar	Large rod
<i>Rps gelatinosa</i>	2.2.1	17011			+	+	+	Binary fission	Tubular	Small rod
<i>Rps capsulata</i>	2.3.1	11166	+			+		Binary fission	Vesicular	Rod
	St. Louis	23782	+			+		Binary fission	Vesicular	Rod
<i>Rps sphaeroides</i>	2.4.1	17023	+			+		Binary fission	Vesicular	Sphere
<i>Rhodomicrobium</i>	3.1.1	17100		+			+	Bud	Lamellar	Stalked oval
<i>vannielli</i>										

Two distinct members of this class of protein present.

Mitochondrial cytochrome <i>c</i>		10	20	30	40	50	60	70	80	90	100
<i>a</i> <i>Equus caballus</i> ²³		AGDVEKGGKIFVQKCAQCHTVEKGG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>b</i> <i>Euglena gracilis</i> ^{24,25}		AGDAERGKGLFESRAQCHSAQKGG	-----	-----	-----	-----	-----	-----	-----	-----	-----
Cytochrome <i>c</i> ₂											
<i>c</i> <i>Rm. varnielii</i> ¹³		AGDPVKGEQVF-KQCKICHQVGP	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>d</i> <i>Rps. viridis</i> ¹³		QDAASGEQVF-KQCLVCHSIGPGA	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>e</i> <i>R. acidophila</i>		AGDPDAGQKVF-LKCAACHKIGPGA	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>f</i> <i>R. molischianum</i> iso-1		ADAPP--PAF-NQCKACHSIDAG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>g</i> <i>R. fulvum</i> iso-1		ADAP--TAF-NQCKACHSIDAG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>h</i> <i>R. molischianum</i> iso-2		ADAP--AGF-TLCKACHSVEAG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>i</i> <i>R. fulvum</i> iso-2		ADAP--PAF-GMCKACHSVEAG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>j</i> <i>R. rubrum</i> ⁸		EGDAAAGEKV-SKKCLACHTFDQGG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>k</i> <i>R. photometricum</i>		AGDAAVGEKIAKAKCTACHDLNKG	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>l</i> <i>Rps. palustris</i> 2.1.6		QDAAKEAVF-KQCMCHRAD	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>m</i> <i>Rps. palustris</i> 2.1.37		QDAKAGEAVF-KQCMCHRAD	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>n</i> <i>Rps. sphaeroides</i>		QEGDPEAGAKAF-NQCGTCHVVDSTGIAGNAKTGPNLYGVVGTAGTQADFGYGEKAGAGLAWDEEHFVQYVQDPTKFLKEYTGDAK	-----	-----	-----	-----	-----	-----	-----	-----	-----
<i>o</i> <i>Rps. capsulata</i> 2.3.1		GDAARKEEF-NKCTCHSIAPDGTETV-KGAKTGPNLYGVVGTAGTPEFK-YKDSI VALGASGFAWTEEDIATYVKDPAFLKEKTDOKK	-----	-----	-----	-----	-----	-----	-----	-----	-----

Fig. 1 Amino acid sequences of Rhodospirillaceae cytochromes *c*₂ aligned with selected mitochondrial cytochrome *c* sequences. Residues are numbered from the horse (*a*) sequence. Underlined residues are referred to in this caption. In sequence *b* residue 86 is ϵ -N-trimethyl-lysine. In sequence *e* the evidence for the amide ascriptions among residues 87–90 is weak. In sequences *h* and *i* position 67 contains phenylalanine (about 30%) as well as the predominant tyrosine. In sequence *i* only the compositions of regions 32–36 and 93–97 are known. In sequence *o* is for *Rps. capsulata* strain 2.3.1; in strain St Louis, residue 75 + 4 is leucine (not threonine) and residue 78 is threonine (not serine). Experimental evidence will be published elsewhere. The one-letter notation used is that recommended by the IUPAC-IUB Commission on biochemical nomenclature. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; M, methionine; N, asparagine, P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

at the position that corresponds to the highly conserved tyrosine-67 of mitochondrial cytochromes *c*. Such polymorphism is well established in higher organisms, where it can occur through population differences¹⁸, heterozygosity, or multiple genes^{19,20}. Only the last of these explanations can apply to the prokaryotes in the present instance. It is remarkable that this heterogeneity is present at the same site in the proteins of two organisms that have otherwise diverged appreciably.

Our results show that the deduction of phylogenetic information from the sequence of homologous bacterial proteins is not straightforward. Studies of the sequences of cytochromes *c*-551 from different *Pseudomonas* species^{21,22} have shown interspecific differences to be much larger than intraspecific differences. Sequence differences between similar species for these cytochromes *c*-551 are as large as those between mitochondrial cytochromes *c* of mammals and insects, although both classes of protein probably have similar redox functions. Our results for the Rhodospirillaceae emphasise this sequence variability more strongly. The predominant low-spin cytochromes *c* from these organisms fall into three distinct sequence classes whose distribution (Table 1) do not correlate with the genera into which the family has been divided on a morphological basis^{5,6}. Even within one of these sequence classes, the cytochrome *c* is more variable between different species of the bacterial family than it is across the whole range of eukaryotic mitochondria. It could be argued that this sequence heterogeneity in the Rhodospirillaceae is a consequence of the extreme age of the family, as organisms with a similar photosynthetic capability could have flourished in the Precambrian epoch before the appearance of oxygenic photosynthesis. Nevertheless, this theory does not predict the lack of correlation of sequence type with morphology nor the occurrence of the similarities observed between the cytochromes of species of Rhodospirillaceae and those of organisms with very different metabolic patterns, for example, *Rps. capsulata* and *P. denitrificans*¹⁵, *Rhodopseudomonas viridis* and mitochondria¹³ (Fig. 1), or *R. tenue* and *Pseudomonas aeruginosa*¹².

Our observations are consistent with the hypothesis that bacterial evolution proceeds both by the assimilation of genes for single functions or for whole metabolic pathways from other organisms as well as by the gradual processes of mutation and selection. We have no information about the frequency of successful gene assimilation or about the timescale for protein evolution and speciation in bacteria. We have no evidence at all

to suggest that either a gene product (such as a cytochrome *c*) or an assemblage of genes (such as *Rhodospirillum rubrum*) should be stable through geological time.

These studies have been supported in part by grants to R.P.A. from the MRC and to M.D.K. from the NIH (GM-18528) and the NSF (BMS-75-13608). We thank G. W. Pettigrew for *R. photometricum* cytochrome *c*₂.

R. P. AMBLER
MARGARET DANIEL
J. HERMOSO

Department of Molecular Biology,
University of Edinburgh,
Mayfield Road,
Edinburgh, UK

T. E. MEYER
R. G. BARTSCH
M. D. KAMEN

Department of Chemistry A-002,
University of California at San Diego,
La Jolla, California 92093

Received 9 January; accepted 5 March 1979.

- Smith, E. L. & Margoliash, E. *Fedn Proc.* **23**, 1243–1247 (1964).
- Dayhoff, M. O. *Atlas of Protein Structure and Sequence* Vol. 5 (National Biomedical Research Fdn, Washington D.C., 1972).
- Goodman, M., Barnabas, J. & Moore, G. W. *Nature* **233**, 604–613 (1971).
- Van Niel, C. B. *Bacteriol. Rev.* **8**, 1–118 (1944).
- Pfennig, N. A. *Rev. Microbiol.* **31**, 275–290 (1977).
- Pfennig, N. and Trüper, H. G. in *Bergey's Manual of Determinative Bacteriology* 8th edn (eds Buchanan, R. E. & Gibbons, N. E.) 24–64 (Williams and Wilkins, Baltimore, 1974).
- Bartsch, R. G. in *The Photosynthetic Bacteria* (eds Clayton, R. K. & Sistrom, W. R.) 249–279 (Plenum, New York, 1978).
- Dus, K., Sletten, K. & Kamen, M. D. *J. biol. Chem.* **243**, 5507–5518 (1968).
- Salemme, F. R., Freer, S. T., Xuong, N. H., Alden, R. A. & Kraut, J., *J. biol. Chem.* **248**, 3910–3921 (1973).
- Hochman, A., Fridberg, I. & Carmeli, C. *Eur. J. Biochem.* **58**, 65–72 (1975).
- Prince, R. C., Baccarini-Melandri, A., Hauska, G. A., Melandri, B. A. & Crofts, A. R. *Biochim. biophys. Acta* **387**, 212–227 (1975).
- Ambler, R. P., Meyer, T. E. & Kamen, M. D. *Nature* **278**, 661–662 (1979).
- Ambler, R. P., Meyer, T. E. & Kamen, M. D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 472–475 (1976).
- Dickerson, R. E., Timkovich, R. & Almasy, R. J., *J. molec. Biol.* **100**, 473–491 (1976).
- Timkovich, R., Dickerson, R. E. & Margoliash, E. *J. biol. Chem.* **251**, 2197–2206 (1976).
- Dus, K., Flattmark, T., DeKlerk, H. & Kamen, M. D. *Biochemistry* **9**, 1984–1990 (1970).
- Schweingruber, M. E., Stewart, J. W. & Sherman, F. *J. molec. Biol.* **118**, 481–496 (1978).
- Petra, P. H., Bradshaw, R. A., Walsh, K. A. & Neurath, H. *Biochemistry* **8**, 2762–2768 (1969).
- Von Ehrenstein, G. *Cold Spring Harbor Symp. quant. Biol.* **31**, 705–714 (1966).
- Hunter, T. & Munro, A. *Nature* **223**, 1270–1272 (1969).
- Ambler, R. P. & Wynn, M. *Biochem. J.* **131**, 485–498 (1973).
- Ambler, R. P. *Biochem. J.* **137**, 3–14 (1974).
- Margoliash, E., Smith, E. L., Kreil, G. & Tuppy, H. *Nature* **192**, 1125–1127 (1961).
- Pettigrew, G. W. *Nature* **241**, 531–533 (1973).
- Lin, D. K., Niece, R. L. & Fitch, W. M. *Nature* **241**, 533–535 (1973).

Anomalies in amino acid sequences of small cytochromes *c* and cytochromes *c'* from two species of purple photosynthetic bacteria

MOST species of the Rhodospirillaceae¹, the purple nonsulphur photosynthetic bacteria, produce large quantities of cytochrome *c*₂ (ref. 2), a protein which is similar in sequence to mitochondrial cytochrome *c*³. Many species also produce the high-spin haem protein cytochrome *c'* (ref. 2). Cytochromes *c'* have been identified in *Rhodospseudomonas gelatinosa* and in *Rhodospirillum tenue*, but in these organisms the predominant low-spin cytochromes *c* do not have properties similar to those found for cytochromes *c*₂ (ref. 2). We have now determined the amino acid sequences of the cytochromes *c'* and of the predominant low-spin cytochromes *c* from the two organisms (Fig. 1).

Comparison of these sequences with those of analogous cytochromes reveals some anomalies (Table 1). The sequences of the two low-spin cytochromes, *Rps gelatinosa* cytochrome *c*-551 and *R. tenue* cytochrome *c*-553 (Fig. 1a) markedly resemble the cytochromes *c*-551 of denitrifying pseudomonads⁴ and of *Azotobacter vinelandii*⁵. For alignment with the *Pseudomonas* and *Azotobacter* sequences, the *Rps gelatinosa* cytochrome requires the postulation of a three-residue insertion immediately after the methionine residue which is presumed to be the sixth iron ligand, while alignment of the *R. tenue* cytochrome requires the postulation of a five-residue insertion shortly before this methionine residue (Fig. 1a). The *R. tenue* cytochrome *c*-553 is particularly similar in sequence to the cytochrome *c*-551 of *Pseudomonas aeruginosa* (Fig. 1a), even in regions that are variable among the *Pseudomonas* species of known sequence^{4,5}.

The sequences of the cytochromes *c'* (Fig. 1b) can be aligned with each other if a one- and a three-residue deletion are postulated in the *Rps gelatinosa* sequence, and a single one-residue deletion is postulated in the *R. tenue* sequence. The sequences show 31% identity overall when so aligned, with poor matching between residues 16 and 50 (Fig. 1b). This compares with 37% identity for the cytochromes *c*-551 (Fig. 1a) and the

27% determined in another study⁶ for their high potential iron sulphur proteins (HiPIP) (Fig. 1c). The sequences of three other cytochromes *c'* are already known, those of an *Alcaligenes* species⁷, *Rhodospirillum rubrum*⁸ and *Chromatium vinosum*⁹. If a two-residue insertion is postulated, the *Rps gelatinosa* cytochrome *c'* matches well with the *Alcaligenes* protein (48% identity) and if a one-residue insertion is postulated, the *R. tenue* cytochrome *c'* matches well with the *C. vinosum* protein (47% identity). In both these cases, good matching extends throughout the sequences (Fig. 1b). The sequence of the cytochrome *c'* from *R. rubrum* is known⁸ and those from *R. molischianum*, *R. fulvum*, *Rps capsulata* and *Rps palustris* are being investigated. None of these proteins resemble the *R. tenue* protein as closely as that protein resembles the *C. vinosum* cytochrome *c'*; neither do any of them resemble the *Rps gelatinosa* protein as closely as that protein resembles the *Alcaligenes* cytochrome *c'* (R.P.A., T.E.M., R. G. Bartsch and M.D.K., unpublished results). The cytochromes *c*-551 and the cytochromes *c'* of *Rps gelatinosa* and *R. tenue* all seem to be anomalously closer in sequence to proteins from organisms of other families, including non-photosynthetic ones, than they are to proteins of other members of the Rhodospirillaceae. *Rps gelatinosa* and the strain of *Alcaligenes* are very different both physiologically and in electron transport protein complement¹⁰. The pattern of similarity observed for the cytochromes *c'* does not apply to the HiPIPs, as the *R. tenue* HiPIP is more different to that from *C. vinosum* than to any of the others of known sequence⁷. *C. vinosum* has been reported to contain a cytochrome *c*-551 (ref. 11), but the amount present is small and nothing is yet known about its structure. Pfennig¹² has recently discussed the relationship between the Chromatiaceae and the Rhodospirillaceae. A reassessment of the evidence of sulphur metabolism, and a consideration of the genus *Ectothiorhodospira* has lessened the distinction between these families; some similarity between their electron transport proteins can now be expected.

To detect evolutionary anomalies, it has been proposed (for example, by Wilson *et al.*¹³) that the amino acid sequences of several different gene products from a set of organisms should be compared, so that evolutionary trees might be constructed and tested for congruence. Although the comparison of three sets of proteins from three different species does not provide sufficient data on which to base firm conclusions, phylogenetic trees constructed from the results in Table 1 could never be interpreted as congruent, nor would they have any predictive value.

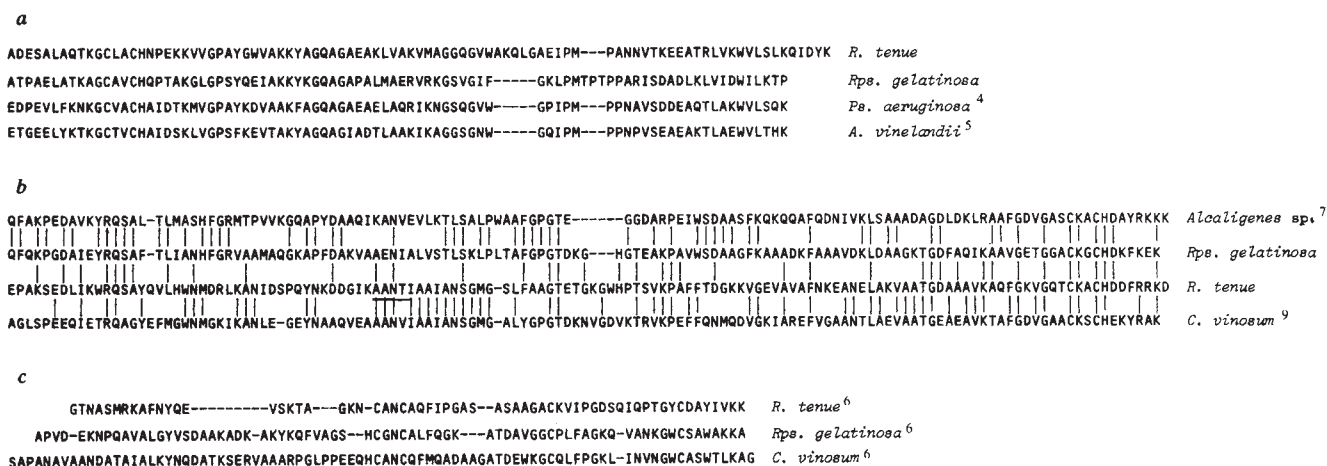


Fig. 1 Amino acid sequences of *a*, cytochromes *c*-551; *b*, cytochromes *c'*; *c*, high potential iron sulphur proteins (HiPIP)⁷ from *Rhodospseudomonas gelatinosa* strain 2.2.1 (= ATCC 17011) and from *Rhodospirillum tenue* strain no. 3761. The sequences are shown aligned with other proteins of the same class and known sequence. In *b*, vertical lines mark identical residues in adjoining sequences. The evidence for the sequence of the underlined residues in *R. tenue* cytochrome *c'* (AANT) is weak. Experimental evidence for the sequences will be published elsewhere. The one-letter notation used is that recommended by the IUPAC-IUB Commission on biochemical nomenclature. A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

Table 1 Similarity matrices for amino acid sequences of cytochromes 'c-551', cytochromes 'c' and high potential iron sulphur proteins (HiPIP)⁶

Cytochromes 'c-551'				Cytochromes 'c'			
<i>Rhodospirillum tenue</i>				<i>Rhodospirillum tenue</i>			
37				31			
56							
45	38	60		47	28		
				29	48	24	
<i>Rhodospirillum tenue</i>				<i>Rhodospirillum tenue</i>			
27							
21	37						
<i>Rhodopseudomonas gelatinosa</i>				<i>Rhodopseudomonas gelatinosa</i>			
56				31			
56	39						
45	38	60		47	28		
				29	48	24	
<i>Pseudomonas aeruginosa</i>				<i>Pseudomonas aeruginosa</i>			
56				31			
56	39						
45	38	60		47	28		
				29	48	24	
<i>Azotobacter vinelandii</i>				<i>Chromatium vinosum</i>			
45				47			
45	38	60		29	48	24	
<i>Alcaligenes sp.</i>				<i>Alcaligenes sp.</i>			
45				29			
45	38	60					

Values shown are matches per 100 residues when the sequences are aligned as in Fig. 1.

Noncongruence of evolutionary trees for *Desulfovibrio* and *Clostridium* ferredoxin and rubredoxin has already been observed¹⁴.

Noncongruence of protein evolutionary trees could be caused by several mechanisms. In higher organisms several anomalies have been explained by the realisation that 'paralogous' proteins had been compared¹⁵. Noncongruence could also be caused if the rate of change of the proteins in separate branches were very different. If functional requirements severely limit the possibility of further change in a protein, the possibility of evolutionary convergence would be increased and would cause further distortion. Nevertheless the simplest (if most difficult to prove) explanation for all the anomalies described is that the proteins concerned have reached some of the organisms by processes of interspecific gene transfer.

These studies have been supported in part by grants to M.D.K. from the NSF (BMS-75-13608) and the NIH (GM-18528) and to R.P.A. from the MRC.

R. P. AMBLER

Department of Molecular Biology,
University of Edinburgh,
Mayfield Road,
Edinburgh, UK

T. E. MEYER
M. D. KAMEN

Department of Chemistry A-002,
University of California at San Diego,
La Jolla,
California 92093

Received 9 January; accepted 5 March 1979.

1. Pfennig, N. & Trüper, H. G. in *Bergey's Manual of Determinative Bacteriology* 8th edn (eds Buchanan, R. E. & Gibbons, N. E.) 24-64 (Williams and Wilkins, Baltimore, 1974).
2. Bartsch, R. G. in *The Photosynthetic Bacteria* (eds Clayton, R. K. & Sistrom, W. R.) 249-279 (Plenum, New York, 1978).
3. Ambler, R. P. *et al.* *Nature* **278**, 659-660 (1979).
4. Ambler, R. P. & Wynn, M. *Biochem. J.* **131**, 485-498 (1973).
5. Ambler, R. P. *Syst. Zool.* **22**, 554-565 (1973).
6. Tedro, S. M., Meyer, T. E. & Kamen, M. D. *J. Biol. Chem.* (in the press).
7. Ambler, R. P. *Biochem. J.* **135**, 751-758 (1973).
8. Meyer, T. E., Ambler, R. P., Bartsch, R. G. & Kamen, M. D. *J. Biol. Chem.* **250**, 8416-8421 (1975).
9. Ambler, R. P., Daniel, M., Meyer, T. E., Bartsch, R. G. & Kamen, M. D. *Biochem. J.* **177**, 819-823 (1979).
10. Suzuki, H. & Iwasaki, H. *J. Biochem.* **52**, 193-199 (1962).
11. Kamen, M. D., Dus, K., Flatmark, T. & DeKlerk, H. in *Electron and Coupled Energy Transfer in Biological Systems* Vol. 1 (eds King, T. E. & Klingenberg, M.) 243-324 (Marcel Dekker, New York, 1972).
12. Pfennig, N. A. *Rev. Microbiol.* **31**, 275-290 (1977).
13. Wilson, A. C., Carlson, S. S. & White, T. J. A. *Rev. Biochem.* **46**, 573-569 (1977).
14. Bruschi, M. & LeGall, J. in *Evolution of Protein Molecules* (eds Matsubara, H. & Yamanaka, T.) 221-232 (Japan. Sci. Soc. Press, 1978).
15. Fitch, W. M. & Margoliash, E. *Evol. Biol.* **4**, 67-109 (1970).

Processing of small subunit precursor of ribulose biphosphate carboxylase and its assembly into whole enzyme are stromal events

CHLOROPLAST development requires the activities of both nuclear and chloroplast genetic systems, but the precise contribution of each and the mechanism by which they interact are not understood. A prime goal is to determine the site of synthesis of each chloroplast polypeptide. Studies with isolated subcellular systems, as well as inhibitor experiments with intact cells, have established that many chloroplast polypeptides are made by cytoplasmic ribosomes¹. Such polypeptides include subunits of both thylakoid and stromal multi-subunit proteins. These observations raise two questions: how do these polypeptides traverse the chloroplast envelope, and where are they assembled into whole functional proteins? The small subunit of the chloroplast enzyme ribulose biphosphate carboxylase is made in the cytoplasm as a higher molecular weight polypeptide precursor². We report here that the cleavage of this precursor and the assembly of the small subunit into whole ribulose biphosphate carboxylase molecules takes place in the chloroplast stroma.

Isolated pea chloroplasts synthesise as the major soluble product the large subunit of ribulose biphosphate carboxylase/oxygenase, the key enzyme in photosynthesis and photorespiration³. The carboxylase enzyme (RBPCase) is an oligomer of 16 subunits. Besides eight large catalytic subunits, the whole enzyme contains eight small subunits (molecular weight 14,000) of unknown function, which are encoded in the nuclear genome⁴. We have shown that the small subunit is synthesised as a precursor of higher molecular weight (20,000) when poly(A)-containing RNA from pea leaf cytoplasmic ribosomes is translated by a heterologous protein-synthesising system^{2,5}. This precursor enters intact isolated chloroplasts with processing to the final size. This uptake and processing does not require concomitant protein synthesis, and thus differs from the signal mechanism advanced to explain the transport of polypeptides across the endoplasmic reticulum⁶. We have suggested an envelope carrier mechanism in which the small subunit precursor is released from cytoplasmic ribosomes and interacts through some aspects of its tertiary structure with a specific binding site in the chloroplast envelope^{2,3}. Unlike known signal sequences, the extra amino acid sequence present in the precursor was postulated to be charged, and its removal by an envelope protease was suggested to trigger a conformational change leading to the transport of the small subunit across the envelope. However, attempts to detect processing activity in preparations of pea chloroplast envelopes purified by the method of Joy and Ellis⁷ have been unsuccessful; instead we find the processing activity to be stromal⁸.

Small subunit precursor (P20) was synthesised by programming a wheat-germ protein-synthesising system with poly (A)-containing RNA from polysomes isolated from etiolated shoots of *Pisum sativum* greened for 48 h (ref. 2). After incubation for 60 min with ³⁵S-methionine the wheat-germ extract was centrifuged at 180,000g for 60 min to remove ribosomes and ribosome-bound labelled polypeptides. The supernatant solution contained P20 as the most heavily labelled polypeptide. Aliquots of this solution were incubated with fractions prepared from intact chloroplasts. Table 1 shows that the processing activity is present in both intact chloroplasts purified on silica-sol gradients, and in stromal extracts prepared from intact chloroplasts. In contrast to previous results², the activity in stromal extracts is not sedimented by centrifugation at 180,000g for 3 h; this treatment sediments both chloroplast ribosomes and whole RBPCase molecules. This soluble stromal nature of the processing activity has been observed in at least 10 separate

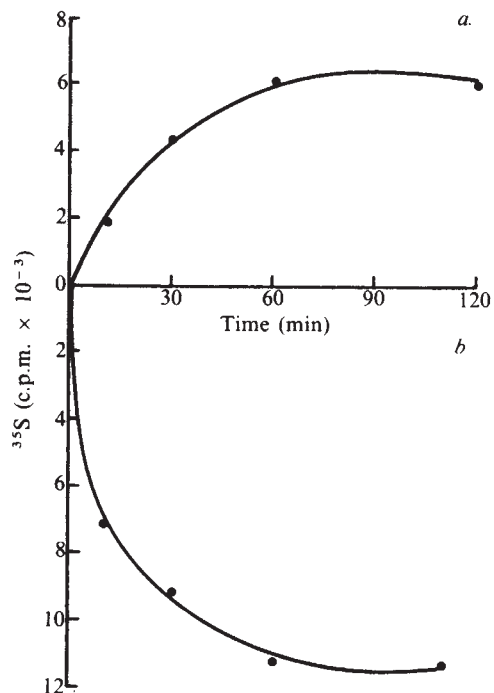


Fig. 1 Time course of processing of P20 to small subunit. Wheat-germ products and 3,200-g stromal supernatant fractions were prepared as in Table 1. Samples (10 μ l) of wheat-germ supernatant were incubated with 40- μ l samples of stroma for various times at 27°C, and then boiled for 2 min in polyacrylamide gel loading buffer². Each incubation mixture contained the equivalent of 5.7 μ g chlorophyll and TCA-insoluble radioactivity equal to 2.5×10^5 c.p.m. Autoradiography and the quantification of the processing activity were as described in Table 1. The results presented are those obtained by counting the radioactivity in the P20 and small subunit bands of the polyacrylamide gel. *a*, Increase in radioactivity in small subunit over that present at zero time; *b*, decrease in radioactivity in P20 below that present at zero time.

experiments carried out independently by two workers. The failure of previous experiments to show that the processing activity is soluble may indicate a loose association of the processing activity with the inner membrane of the chloroplast envelope. The small subunit precursor from *Chlamydomonas* is also processed to the small subunit by a soluble enzyme⁹, but in that case it was not possible to show that the activity was located in the stroma because intact chloroplasts cannot be isolated from this alga.

Loss of label from P20 is not invariably associated with the appearance of label in the small subunit, as we have previously assumed. Membrane preparations often cause a decrease in P20 label, but this loss is never accompanied by an increase in small subunit label (Table 1) or other polypeptide. No other polypeptide synthesised by the wheat-germ system shows a decrease in label after incubation with any chloroplast fraction, suggesting that nonspecific proteolysis is absent. The time course of processing by stromal extracts shows that more label is lost from P20 than appears in the small subunit (Fig. 1). This observation establishes that the sequence removed during processing contains at least one methionine residue. If this methionine residue is at the N terminus, its removal by an N-terminal peptidase would account for the activity seen in the membrane preparations. Such N-terminal peptidase activity may also occur in the stromal fractions, but this is impossible to establish until specific inhibitors of processing are available.

The soluble nature of the P20 processing activity raises the question as to how it functions in the transport of the precursor through the chloroplast envelope. One possibility is that the precursor interacts with the envelope in such a way that the

sequence to be removed becomes accessible from the stromal side. Removal of the sequence may be the event which triggers the release of small subunit into the stromal compartment. A critical test of this idea would be to determine whether removal of the extra sequence by purified processing enzyme affects the ability of intact chloroplasts to take up the molecule.

An additional or even alternative role for processing may be in the assembly of whole RBPCase molecules. Analysis on

Table 1 Subcellular location of small subunit precursor processing activity

Fraction tested	% Decrease in label in small subunit precursor (P20)	% Increase in label in small subunit
Buffer	0	0
Washed chloroplasts	41	75
Purified intact chloroplasts	31	50
Lysed chloroplasts	33	45
Stroma 3,200g 5 min	30	58
Stroma 180,000g 1 h	41	47
Stroma 180,000g 3 h	40	62
Thylakoids	21	0
Envelopes	35	0

Poly(A)-containing RNA was prepared from pea leaf polysomes and translated in a wheat-germ protein-synthesising system as before², except that the final concentrations of KCl and MgCl₂ were 90 mM and 2 mM, respectively. After incubation, the wheat-germ extract was centrifuged at 180,000g for 60 min at 4°C to sediment ribosomes and ribosome-bound polypeptides. This treatment sediments about 60% of the trichloroacetic acid (TCA)-insoluble radioactivity, leaving the small subunit precursor in the supernatant as the most heavily labelled polypeptide. During the preparation of the wheat-germ products chloroplasts were isolated from 10-d-old pea leaves as described previously³. These chloroplasts were washed once in isolation medium, and resuspended in intact plastid buffer prepared by mixing equal volumes of 10 mM Tris-HCl, pH 7.6 and 50 mM HEPES-KOH, 220 mM KCl, 6 mM MgCl₂, 20 mM dithiothreitol (DTT), pH 7.6. This preparation is referred to as washed chloroplasts, and consists of about 60% intact chloroplasts as judged by phase-contrast microscopy. Lysed chloroplasts were prepared by resuspending a pellet of washed chloroplasts in 10 mM Tris-HCl, pH 7.6, leaving it on ice for 5 min, and then adding an equal volume of 50 mM HEPES-KOH, 220 mM KCl, 6 mM MgCl₂, 20 mM DTT, pH 7.6. A sample of lysed chloroplasts was centrifuged at 3,200g for 5 min to obtain a stromal supernatant and a thylakoid pellet. The latter was washed once, and resuspended in an equal volume of intact plastid buffer. Samples of stroma were centrifuged further at 180,000g for 1 and 3 h at 4°C while other preparations remained on ice. Intact chloroplasts were purified on gradients of Ludox AM silica sol¹² from a preparation of washed chloroplasts. Purified envelopes were prepared from washed chloroplasts by the method of Joy and Ellis⁷. For the measurement of processing activity, samples (10 μ l) of wheat-germ supernatant containing TCA-insoluble radioactivity equal to 2×10^5 c.p.m. were incubated with 40 μ l of chloroplast preparation, or fractions therefrom, for 60 min at 27°C. Chlorophyll was measured by the procedure of Arnon¹³; each incubation contained the equivalent of 6 μ g chlorophyll, except that the purified chloroplast envelope incubation mixtures contained envelopes equivalent to 990 μ g chlorophyll. After incubation, samples were prepared for polyacrylamide gel electrophoresis and analysed as before². Gels were autoradiographed with Kodirex X-ray film for 100 h. Processing activity was measured quantitatively by two methods. Autoradiographs were scanned in a Joyce-Loebl densitometer, and the peak heights of the P20 and small subunit bands were measured for incubated and non-incubated samples. The change in peak height of each band due to the incubation was calculated as a percentage. In the other method the dried polyacrylamide gels were rehydrated, and the P20 and small subunit bands were excised. The gel pieces were dissolved in 0.2 ml H₂O₂ (100 volumes) at 50°C overnight, and the radioactivity in each was counted in 4 ml Triton-toluene scintillant (0.4% (w/v) PPO, 0.005% (w/v) POPOP, 33% (v/v) Triton X-100 in toluene). The results presented were obtained by the first method and agree closely with those obtained by the second method. The percentage increase in small subunit label is not correlated with the percentage decrease in P20 label; this is because both the P20 and small subunit regions of the gel are contaminated with other products of the wheat-germ system in both the presence and absence of processing.

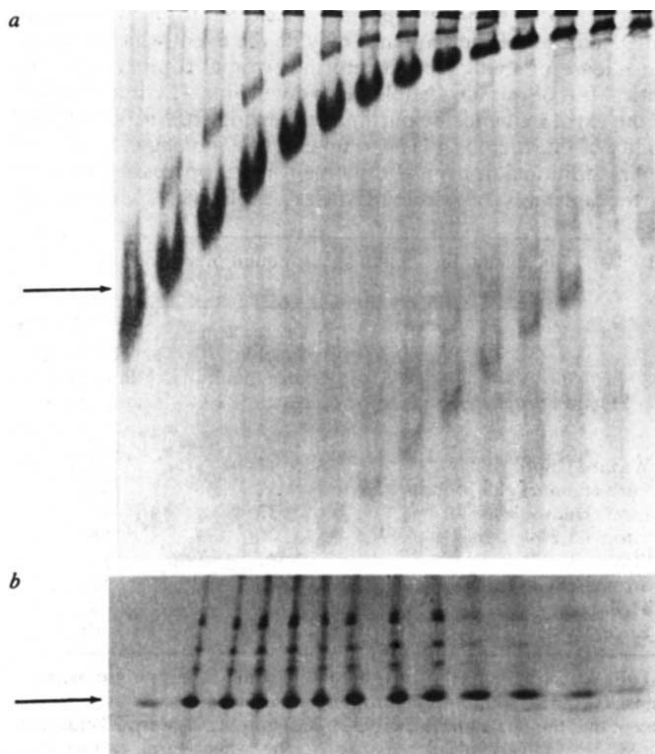


Fig. 2 Assembly of small subunit into whole RBPCase molecules. *a*, Separation of whole RBPCase molecules by transverse gradient non-denaturing gel electrophoresis. Wheat-germ products and a 3,200g stromal supernatant were prepared as in Table 1. Wheat-germ supernatant (50 μ l) containing 1.25×10^6 c.p.m. of TCA-insoluble radioactivity was incubated with 200 μ l of stromal extract, equivalent to 28.8 μ g chlorophyll, for 60 min at 27 $^{\circ}$ C. A control incubation mixture consisted of 50 μ l wheat-germ supernatant plus 200 μ l intact plastid buffer. After incubation, samples were brought to 300 μ l with a solution of sucrose and 2-mercaptoethanol with final concentrations of 10% (w/v) and 1.66% (v/v) respectively, for analysis by non-denaturing polyacrylamide gel electrophoresis. The gel system is based on that of Blair and Ellis³. A slab gel (17 $\times$ 17 cm) was cast with a linear gradient of 5–10% (w/v) acrylamide and 0.1–0.2% (w/v) bisacrylamide, stabilised with 0.2% (w/v) linear polyacrylamide. Once polymerised, the gel was turned through 90 $^{\circ}$, and a 5% (w/v) acrylamide, 0.1% (w/v) bisacrylamide slab former was cast along one edge of the gradient gel. The gel was pre-electrophoresed for 3.5 h at 30 mA. Samples of incubation mixture containing 2×10^5 c.p.m. TCA-insoluble radioactivity were loaded onto all the slots. Electrophoresis was for 18 h at 27 mA after which the gel was stained with Coomassie brilliant blue R and destained². The photograph shows the stained gel. RBPCase is the most heavily stained band, and is marked with an arrow at the 5% end of the gel. Other, minor bands, change their mobility relative to RBPCase according to the gel concentration. Autoradiography for 100 h shows radioactivity co-migrating with the stained bands of RBPCase, but not at corresponding positions in tracks lacking stromal extract (not shown). *b*, Identification of labelled small subunit in whole RBPCase molecules. The transverse gradient non-denaturing polyacrylamide gel shown in (*a*) was rehydrated, and each RBPCase region was excised. Gel pieces were equilibrated twice with 5 ml of 62 mM Tris-HCl, 2% (w/v) SDS, 0.5% (v/v) 2-mercaptoethanol pH 6.8 for 15 min each, and then affixed to the top of an SDS slab gel with 1% (w/v) Agarose, 0.2% (w/v) SDS, 62 mM Tris-HCl, pH 6.8. Electrophoresis was as previously described². The photograph shows an autoradiograph of the SDS gel in the small subunit region; the arrow marks the small subunit marker. The autoradiograph shows that radioactivity co-migrating with small subunit is present in RBPCase molecules excised from across the entire gradient of the non-denaturing gel. Other radioactive polypeptides migrate above and below the small subunit. However, these polypeptides vary in distribution across the gradient, indicating that they do not co-migrate with RBPCase at all gel concentrations. Control incubations lacking stromal extract contain no radioactivity in the small subunit region (not shown).

non-denaturing gels shows that the large subunit of RBPCase synthesised by isolated pea chloroplasts does not enter pre-existing RBPCase molecules³, but accumulates as an aggregate¹. This observation has been interpreted as indicating that isolated pea chloroplasts lack a pool of small subunits³. It follows that supplying P20 to pea chloroplast fractions may result in the assembly of whole RBPCase molecules. When wheat-germ incubation mixture containing labelled P20 is incubated with a stromal extract, labelled small subunit co-migrates with whole RBPCase molecules over a wide range of polyacrylamide gel concentrations when electrophoresed in non-denaturing conditions (Fig. 2). Co-migration in such conditions is a good test for identity¹⁰. Control experiments where stromal extracts were omitted from the incubation show no label migrating with whole RBPCase molecules. These observations establish that the assembly of small subunit into RBPCase occurs in the chloroplast stroma as previously proposed², and not in the cytoplasmic compartment. Chua and Schmidt¹¹ have reported the processing of P20 and assembly of small subunit into RBPCase in intact pea and spinach chloroplasts, but the stromal location of the processing and assembly process was not established. Whether processing is an integral part of the assembly mechanism, or merely provides small subunit for that mechanism, can be resolved when the processing activity has been purified.

This work was supported by the SRC.

STEVEN M. SMITH

R. JOHN ELLIS

Department of Biological Sciences,
University of Warwick,
Coventry, UK

Received 18 January; accepted 13 March 1979.

1. Ellis, R. J. *Biochim. biophys. Acta* **463**, 185–215 (1977).
2. Highfield, P. E. & Ellis, R. J. *Nature* **271**, 420–424 (1978).
3. Blair, G. E. & Ellis, R. J. *Biochim. biophys. Acta* **319**, 223–234 (1973).
4. Kung, S. D. *Science* **191**, 429–434 (1976).
5. Ellis, R. J., Highfield, P. E. & Silverthorne, J. *Proc. Symp. Fourth int. Congr. Photosynthesis* (eds Hall, D. O., Coombs, J. & Goodwin, T. W.) 497–506 (Biochemical Society, London, 1977).
6. Blobel, G. & Dobberstein, B. *J. cell. Biol.* **67**, 835–851 (1975).
7. Joy, K. W. & Ellis, R. J. *Biochim. biophys. Acta* **378**, 143–151 (1975).
8. Ellis, R. J. & Barraclough, B. R. *Chloroplast Development* (ed. Akoyunoglou, G.) 185–194 (North-Holland, Amsterdam, 1978).
9. Dobberstein, B., Blobel, G. & Chua, N.-H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1082–1085 (1977).
10. Hedrick, J. L. & Smith, A. J. *Archs. Biochem. Biophys.* **126**, 155–164 (1968).
11. Chua, N.-H. & Schmidt, G. W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6110–6114 (1978).
12. Morgenthaler, J.-J., Marsden, M. P. F. & Price, C. A. *Archs Biochem. Biophys.* **168**, 289–301 (1975).
13. Arnon, D. I. *Pl. Physiol.* **24**, 1–15 (1949).

Sites of inhibition of *in vitro* DNA synthesis in carcinogen- and UV-treated Φ X174 DNA

THE availability of DNA molecules of known sequence and rapid methods for the determination of the sequence of particular DNA fragments^{1–3} make it possible to investigate the molecular action of mutagens and carcinogens more closely^{4,5}. For example, there is considerable evidence that some types of lesion in DNA such as UV light-induced pyrimidine dimers and various chemical adducts are blocks to DNA synthesis both *in vivo*^{6–9} and *in vitro*^{10,11}. The available data support the hypothesis that the block occurs at the level of the lesion on the template strand^{10,11}. We have used DNA containing such lesions on a template of known sequence for *in vitro* DNA synthesis by DNA polymerase 1 to investigate the site of the block at the level of the nucleotide sequence. This method allows us to

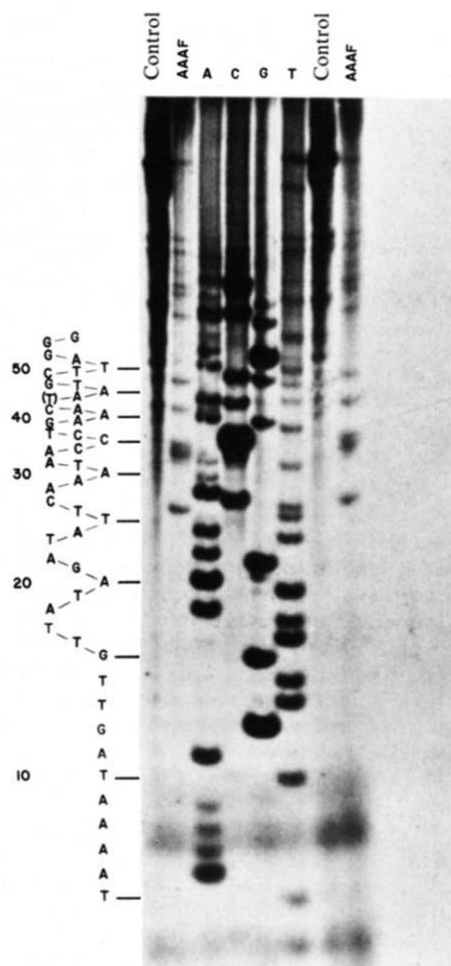


Fig. 1 Polyacrylamide gel analysis of the product synthesized by polI on a DNA template reacted with AAAF. Single-stranded phage Φ X174 DNA ($400 \mu\text{g ml}^{-1}$) was reacted with ^{14}C -AAAF ($50 \mu\text{g ml}^{-1}$) for 4 h at 30°C , extracted 4 times with amyl alcohol, precipitated with ethanol and resuspended in 10 mM Tris-HCl pH 7.5, 0.25 mM EDTA. The extent of reaction, determined from the ratio of ^{14}C c.p.m. to A_{260} was ~ 265 AAAF adducts per Φ X DNA molecule. Double-stranded RF Φ X DNA was digested with *Hae*II endonuclease (BRL) and individual restriction fragments were purified on agarose gels. Restriction fragment 2 (ref. 12) was heated at 100°C for 5 min and annealed to either AAAF-treated or untreated DNA at 65°C for 90 min (molar ratio 1.2:1). Polymerase reactions were carried out in a volume of $20 \mu\text{l}$ and contained 50 mM Tris-HCl, pH 8.0, 8 mM MgCl_2 , 5 mM dithiothreitol, dATP, dGTP and dCTP, $50 \mu\text{M}$ each, $1\text{--}2 \mu\text{Ci}$ [$\alpha\text{-}^{32}\text{P}$]TTP (about 300 Ci mmol^{-1}), $0.1\text{--}0.2 \mu\text{g}$ DNA and 0.7 units DNA polymerase I (Klenow fragment, Boehringer). Incubation was at room temperature for 15 min and then $2 \mu\text{l}$ 0.5 mM TTP was added, incubation continued for a further 15 min and the reaction was stopped by heating at 65°C for 10 min. One unit of *Hae*II (BRL) was then added and incubated at 37°C for 60 min. For the sequence standards di-deoxyribonucleoside triphosphates were used as described by Sanger *et al.*³. The products were denatured and analysed on a 20% polyacrylamide gel². Lanes Control and AAAF are the products synthesized on untreated and AAAF-reacted templates respectively. Lanes A, C, G and T are the sequence standards produced with the di-deoxy chain terminating nucleotides. Numbering is from the position of the *Hae*II cleavage site.

determine both the site at which lesions capable of blocking synthesis by DNA polymerase occur and exactly where synthesis stops in relation to the lesion.

Φ X174 DNA was used in these experiments as it has already been completely sequenced^{12,13}. The Klenow fragment of *Escherichia coli* polI was used as the polymerase for most of our

experiments. The template consisted of single-stranded phage DNA to which individual restriction fragments, obtained by digestion of the double-stranded RF DNA, were annealed. We used fragments produced by the enzyme *Hae*II which cuts Φ X174 RF DNA at eight sites¹². These templates were used for the di-deoxynucleotide chain terminating method of Sanger and Coulson to provide a sequence standard. We found minor discrepancies with the preliminary sequence published by Sanger *et al.*¹²; however, in practically every case these were found to be in complete agreement with the revised sequence¹³. Single-stranded Φ X174 DNA was either UV-irradiated or reacted with *N*-acetoxy-2-acetyl aminofluorene (AAAF) or other carcinogens. The reacted template was then annealed with the priming restriction fragment and the polymerisation reaction carried out as for the sequencing reactions except that the di-deoxynucleotides were omitted.

In the reaction with AAAF-treated DNA a number of bands are seen which do not appear when untreated DNA is used (Fig. 1). These bands in the three cases in which identification is

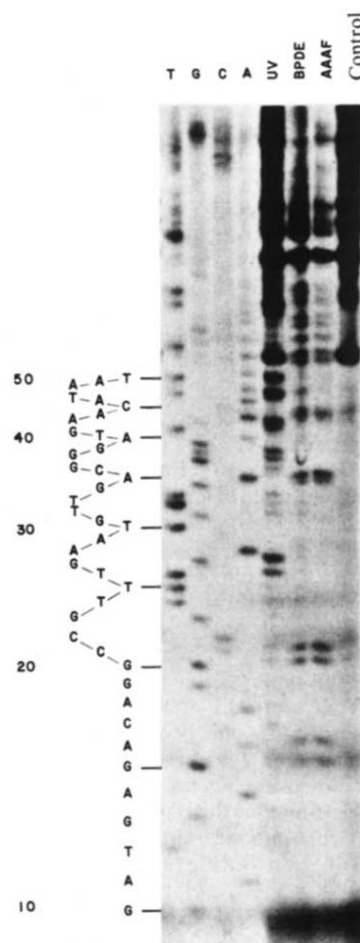


Fig. 2 Polyacrylamide gel analysis of the products synthesized by polI on DNA templates reacted with AAAF or anti-BPDE or irradiated with UV light. For the anti-BPDE template $46 \mu\text{g}$ Φ X DNA was incubated at 37°C with 30 nmol ^{14}C -anti-BPDE in 0.5 ml . At 60-min intervals a further three additions of ^{14}C -anti-BPDE were made and incubation continued for 5 h. The DNA was extracted four times with amyl alcohol and precipitated with ethanol. About 140 anti-BPDE adducts were added per Φ X DNA molecule. The DNA for the UV template was irradiated with 1000 J m^{-2} of UV at a DNA concentration of $100 \mu\text{g ml}^{-1}$. The primer used in this experiment was *Hae*II fragment 5. Reaction conditions were the same as for Fig. 1. Lanes Control, AAAF, BPDE and UV are the products synthesized by polI on untreated, AAAF or anti-BPDE-reacted and UV-irradiated DNA templates respectively. Lanes A, C, G and T are the sequence standards.

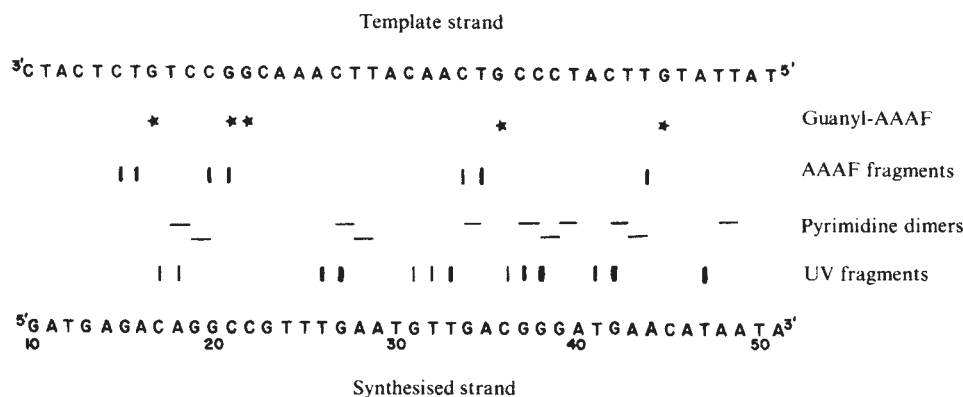


Fig. 3 The position of chain termination in relation to the sites of lesions in the DNA: diagrammatic representation of data shown in Fig. 2. Potential sites for the formation of *N*-acetyl-*N*-(guan-8-yl)-2-aminofluorene (guanines) and of pyrimidine dimers (adjacent pairs of pyrimidines) are shown along side of the position of bands obtained with synthesis by pol1 on AAAF-reacted or UV-irradiated DNA templates.

possible occur one base before that of a cytosine residue in the sequence standards, that is, synthesis has stopped at the base before a guanine in the DNA template. AAAF has been shown to react predominantly with guanine in DNA to form *N*-acetyl-*N*-(guan-8-yl)-2-aminofluorene^{14,15}. Our results both confirm the known major site of reaction of AAAF in DNA and demonstrate that this lesion is a block to DNA synthesis. Another conclusion that can be drawn from these experiments is that there seems to be no preferential reaction of AAAF with particular guanines in the section of tested DNA, at least in the conditions used here; that is, single-stranded DNA reacted *in vitro*. Bands can be detected corresponding to all the guanines in the region of the template for which the sequence can be resolved, and there is no obvious predominance of any of them. This may simply reflect the very high number of adducts in the DNA (about 1 per 5 guanines) although with a lower number of adducts (about 1 per 30 guanines) no bands at all were detected. There is no indication of lesions blocking synthesis at any position other than guanine. This conclusion was confirmed by an experiment using a different priming fragment (Fig. 2) which included templates either treated with ($\pm$)7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydro benzo(a)pyrene (anti-BPDE) or irradiated with UV light. The pattern of bands obtained in this experiment is shown in Fig. 3. The bands produced with the AAAF-treated template are again consistent with those expected if synthesis terminated one nucleotide before an *N*-acetyl-*N*-(guan-8-yl)-2-aminofluorene residue. A completely different pattern of bands is seen with the UV-irradiated template. Synthesis on such a template terminates one nucleotide before any pair of pyrimidine bases. All three types of pyrimidine dimer, cytosine-cytosine, cytosine-thymine and thymine-thymine, are blocks to DNA synthesis although there is some indication that the bands resulting from thymine-thymine dimers are more prominent, probably due to their faster rate of formation during irradiation¹⁶ and thus more frequent occurrence in the template.

The results with anti-BPDE are more difficult to interpret. Prominent bands occur at the same positions as found with AAAF treatment although many other bands can also be seen. This would suggest that guanine is a major site of reaction for anti-BPDE¹⁷ although reaction also occurs at other bases, notably adenine.

As synthesis stops at the base preceding a guanine with AAAF-treated template DNA and at the base before the first of the two pyrimidine bases with UV-treated template DNA, it is clear that termination occurs without a nucleotide present at the site(s) opposite the lesion. It is possible that DNA polymerase 1 does add nucleotides in this position but that they are rapidly excised by the 3'-5' exonuclease activity of the enzyme due to lack of base pairing¹¹; however, our observation (data not shown) that a DNA polymerase α isolated from human lymphoblastoid cells¹⁸ and having no 3'-5' exonuclease activity gives the same band pattern with UV-irradiated template DNA

as does pol1 suggests that synthesis terminates due to the inability of the enzyme to incorporate nucleotides opposite a non-pairing lesion. We occasionally found additional slightly lighter bands one or even two bases before the main termination bands, for example, in Figs 2 and 3 at positions 15 and 34 with AAAF and at positions 31 and 32 with UV. These may result from degradation by either the 3'-5' exonuclease or residual 5'-3' exonuclease that contaminates some commercial batches of pol1 (Klenow). Degradation by the 5'-3' exonuclease is considered unlikely as treatment at 65°C before addition of the restriction enzyme completely eliminated spurious bands on the sequence standards caused by this activity but did not affect the extra bands on the AAAF-reacted or UV-irradiated templates.

These experiments are essentially qualitative and do not show that DNA polymerase stops at every lesion; one cannot determine by this method the amount of 'read through'. However, we expect that the method will be useful to distinguish between lesions that are blocks to DNA synthesis and those that are not as well as elucidating the behaviour of different DNA polymerases on encountering a lesion and possible ways in which DNA synthesis may be able to bypass lesions¹⁹.

We thank Dr Geoffrey Grigg for the conversation during which these experiments were devised. The work was supported by an ACS institutional grant (IN-41-R-30) and from NIH (GM 07816-20).

PETER MOORE

BERNARD S. STRAUSS

Department of Microbiology,
The University of Chicago,
Chicago, Illinois 60637

Received 29 December 1978; accepted 23 February 1979.

- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
- Sanger, F. & Coulson, A. R. *J. molec. Biol.* **94**, 441-448 (1975).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Hatayama, T., Goldberg, I. H., Takeshita, M. & Grollman, A. P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3603-3607 (1978).
- D'Andrea, A. D. & Haseltine, W. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3608-3612 (1978).
- Rupp, W. P. & Howard-Flanders, P. *J. molec. Biol.* **31**, 291-304 (1968).
- Cailliet-Fauquet, P., Defais, M. & Radman, M. *J. molec. Biol.* **117**, 95-112 (1977).
- Lehmann, A. R. *J. molec. Biol.* **66**, 319-337 (1972).
- D'Ambrosio, S. M. & Setlow, R. B. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2396-2400 (1976).
- Hsu, W. T., Lin, E. J. S., Harvey, R. G. & Weiss, S. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3335-3339 (1977).
- Villani, G., Boiteux, S. & Radman, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3037-3041 (1978).
- Sanger, F. *et al. Nature* **265**, 687-695 (1977).
- Sanger, F. *et al. J. molec. Biol.* **125**, 225-246 (1978).
- Kriek, E., Miller, J. A., Juhl, U. & Miller, E. C. *Biochemistry* **6**, 177-182 (1967).
- Kriek, E. *Biochim. biophys. Acta* **355**, 177-203 (1974).
- Setlow, R. B. & Carrier, W. L. *J. molec. Biol.* **17**, 237-254 (1966).
- Meehan, T., Straub, K. & Calvin, M. *Nature* **269**, 725-727 (1977).
- Bose, K., Karran, P. & Strauss, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 794-798 (1978).
- Higgins, N. P., Kato, K. & Strauss, B. *J. molec. Biol.* **101**, 417-425 (1976).

Metastable crystals of β -cyclodextrin complexes and the membrane diffusion model

WE have previously suggested¹ that dimers of β -cyclodextrin (β -CD) may mimic membrane diffusion transport phenomena. This proposal was based on the orientation of guest molecules in the crystalline 3:2 *p*-iodophenol complex with β -CD, with the assumption that complex formation with guest molecules capable of hydrogen bond formation occurs through an initial interaction with one of the hydroxyl groups of β -CD. The 7:14 distribution of hydroxyl groups on the primary and secondary sides of the β -CD torus, respectively, was interpreted as favouring formation of a complex with the hydrogen bonding moiety on the secondary side of the torus. The fact that the guest molecules in the above complex were orientated in the opposite direction was taken as an indication that exchange of guest molecules between a dimeric complex and the solution had occurred. We report here that β -CD inclusion complexes with various guest molecules produce metastable crystals (designated CII) when growth is achieved by slow cooling, and we have determined the crystal structure of an *n*-propanol complex of

this phase. We find that comparison of the structure of this metastable phase and that of the thermodynamically more stable phase¹, CI, provides information relevant to the question of the presence of β -CD dimers in solution; this information is relevant to the suggestion that such dimers mimic bilayer membrane diffusion transport.

CII crystals of the *n*-propanol complex were grown from a 60:40 (v/v) *n*-propanol:water solution of β -CD by slow cooling. Typically, a solution near its boiling point was contained in a beaker sealed with a wax sheet topped with aluminium foil; the beaker was placed in a Dewar flask containing a water bath also near its boiling point. The flask was covered with thermal insulating material and allowed to stand for 1.5–2 d, during which the system cooled to a temperature describable as warm to tepid. Crystals with well formed faces and uniform morphology were routinely obtained in this manner. After several hours at room temperature, these crystals had obviously begun to dissolve and crystals of type CI appeared; their number increased with time as those of type CII decreased until the latter phase had disappeared.

Both crystalline phases were found to display P1 space-group symmetry but to have different unit cell volumes: $V_{CI} = 3,636 \text{ \AA}^3$ and $V_{CII} = 3,346 \text{ \AA}^3$ (both at $\sim 120 \text{ K}$); the lattice parameters for CII are $a = 15.461(2)$, $b = 15.575(3)$, $c = 15.316(3) \text{ \AA}$, $\alpha = 103.98^\circ(2)$, $\beta = 100.92^\circ(2)$ and $\gamma = 104.23^\circ(2)$. The initial crystal structure was solved by the use of

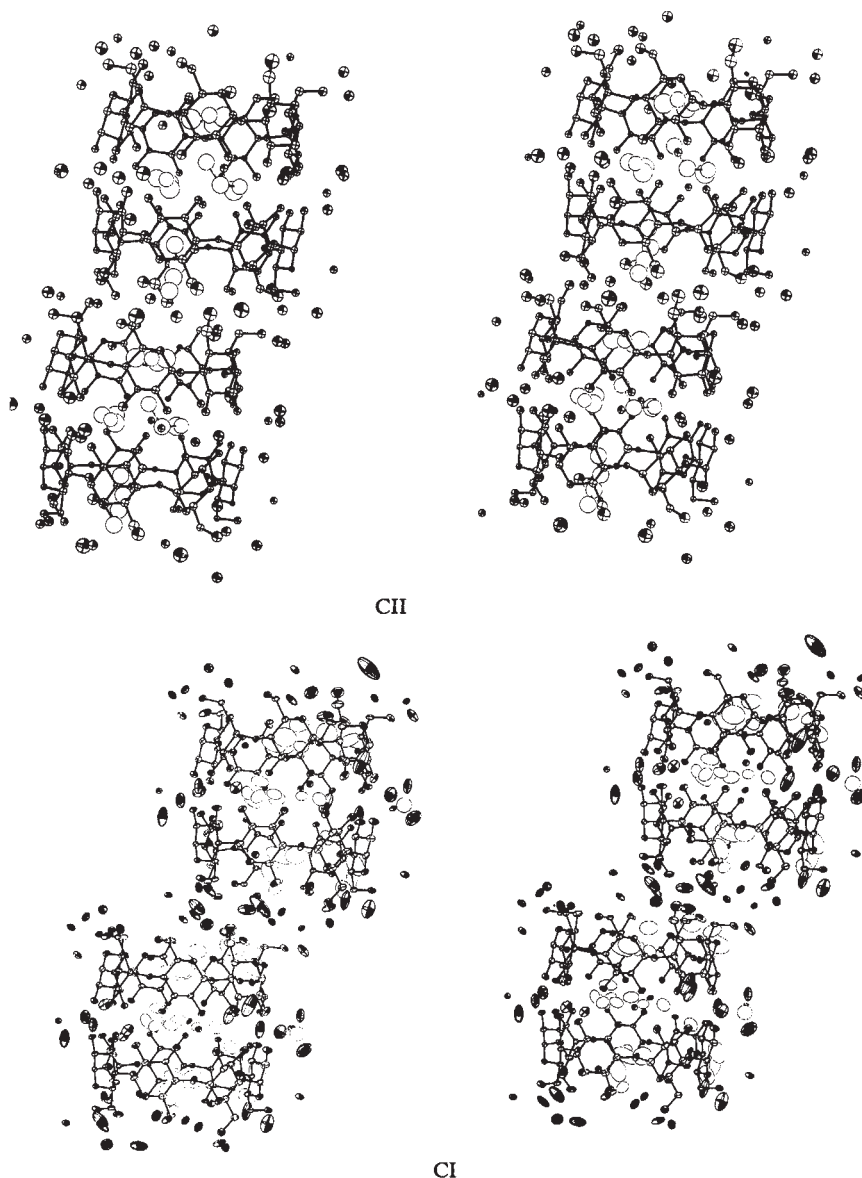


Fig. 1 Stereoscopic projections² of the crystal structure of *n*-propanol complexes of β -CD in thermodynamically stable phase CI and metastable phase CII. To illustrate the packing differences, two β -CD:*n*-propanol dimers have been drawn. The oxygen atoms of their environment (water molecules or adjacent β -CD molecule hydroxyl oxygen atoms) have been drawn. The disorder in the guest molecules is indicated by the size and shape of their thermal parameter representations. Those of phase CI are indicative of anisotropic thermal parameters whereas those of CII correspond to isotropic temperature factors.

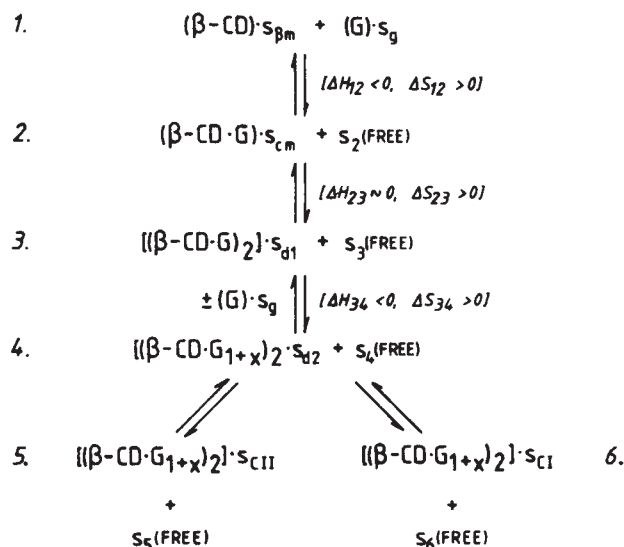


Fig. 2 A schematic representation of the principal processes leading to crystals of phases CI (thermodynamically stable) and CII (metastable) as the result of slow cooling of an aqueous solution. The symbols are: β -CD, β -cyclodextrin; G, guest molecules suitable for formation of inclusion complex β -CD·G, which may be of different stoichiometry in the monomeric and dimeric forms; s, solvation molecules (subscripted to indicate possible differences in solvation for the various species indicated). Trends in thermodynamic properties, ΔH , ΔS , and ΔG , are indicated for changes between numbered states.

non-crystallographic symmetry and the known structure of the β -CD dimer from the crystal structure analyses of phase CI, in a manner that will be described in more detail elsewhere. Refinement of the structural model of CII with isotropic temperature factors and a high resolution data set ($2\theta_{\max} = 60^\circ$, MoK α radiation) has converged at $R = 12.0\%$ for 15,000 contributing data. At this stage of the refinement the unit cell composition is $C_{96}H_{172}O_{74} \cdot 18\text{--}19H_2O$. Refinement of the structural model for the CI *n*-propanol complex with anisotropic temperature factors is nearing completion. We now believe that the complex contains four disordered *n*-propanol guest molecules, giving a unit cell composition of $C_{96}H_{174}O_{74} \cdot 24H_2O$. The present R value is 8.0% for $\sim 19,300$ contributing reflections ($2\theta_{\max} = 60^\circ$).

Stereoscopic projections² displaying two β -CD dimers and oxygen atoms of their environment are presented for the *n*-propanol- β -CD complexes of phases CI and CII in Fig. 1. It is immediately apparent from the figure that the crystal packing is very different in the two phases. The metastable phase has a channel structure that seems to be similar to the packing reported for other β -CD complexes grown from aqueous solution by slow cooling^{3,4}, whereas phase CI displays a brickwork packing network. Another significant difference between the structures of the two phases is the presence of a water layer between the β -CD dimers in CI and its absence from the metastable phase. We find these structural differences and the metastable property of phase CII intriguing because of the inferences they have for the structure and properties of β -CD complexes in concentrated solutions.

The temperature-solubility gradient displayed by the β -CD:*n*-propanol:water system was used to grow crystals of phase CII. The fact that these crystals are metastable implies that they result from a kinetically rather than thermodynamically controlled process. Figure 2 summarises what we believe to be the principal processes giving rise to phase CII and, subsequently, to phase CI.

Complex formation to give β -CD·G, step 12, is expected⁵ to increase the entropy of the system as the result of release of molecules of solvation from both host and guest and to decrease

the enthalpy by freeing high energy (ref. 5) water molecules from the β -CD torus for subsequent hydrogen bond formation. We suggest that solution molecular dynamics are important in step 23. Water molecules have a considerably higher ($\sim 10\times$) mean velocity than the larger β -CD monomers. At temperatures at which thermal dissociation of hydrogen bonds is prevalent, this velocity difference can be important in determining which molecules form stable interactions. Two properties of β -CD·G monomers favour their association: their similar molecular velocities and the favourable distribution of secondary hydroxyl groups for the formation of multiple interdimer hydrogen bonds; dimerisation is also thermodynamically favourable (see Fig. 2). The inclusion of additional guest molecules in the extended torus of the dimer, step 34, is thermodynamically favourable but not crucial to the remainder of the process. We include it here to account for the four *n*-propanol molecules in these structures. Step 45, crystallisation of phase CII, is the kinetically controlled step in the process and we believe it reflects the solution molecular dynamics described for step 23. This proposal is clearly consistent with the nucleation of phase CII at temperatures well above room temperature.

The metastable character of phase CII crystals thus indicates that another component in the equilibrium becomes important at lower temperature. We suggest that this component is the building in solution of a stable hydration shell over the primary faces of the dimer (a component of step 46). This step may be accompanied by exchange of guest molecules with the solution and may also result in a change in 'stoichiometry' of the dimer. Orientation of hydrophilic substituents of the included molecules towards the primary side of the β -CD torus can be expected to stabilise the hydration layer further by promoting the formation of hydrogen bonds. Crystallisation of the more highly hydrated dimer gives rise to an equilibrium system and the thermodynamically stable phase CI.

We interpret the presence of the β -CD·G dimers in phase CII, and the observation that crystals of this phase dissolve at room temperature to give crystals of a second phase also consisting of β -CD·G dimers, as evidence that the predominant species in concentrated aqueous solutions is the dimer; a recent solution study supports this interpretation⁶.

The orientation of the guest molecules in the two structures also differs but it is inappropriate to take this as evidence for exchange of guest molecules with solution, as the *n*-propanol molecules are sufficiently small to allow their reorientation within the torus. The disorder in the *n*-propanol molecules in phase CI makes it unreliable to assign C and O atoms of the guest molecules in the middle of the dimer. However, it is possible that those in the ends of the dimer are hydrogen bonded to hydrate water molecules. There seems to be a short chain of hydrogen bonds involving the guest molecules and one hydrate molecule in phase CII, but we prefer to wait for the results of further refinement before seeking to analyse it further.

We thank the Deutsche Forschungsgemeinschaft (grant Ste 230/2) and the Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart for making this research possible.

KURT H. JOGUN

JOHN J. STEZOWSKI*

Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart, Pfaffenwaldring 55, D-7000 Stuttgart 80, FRG

Received 21 November 1978; accepted 5 March 1979.

* To whom correspondence should be addressed.

1. Stezowski, J. J., Jogun, K. H., Eckle, E. & Bartels, K. *Nature* **274**, 617–619 (1978).
2. Johnson, C. K. *ORTEP II: A FORTRAN Thermal Ellipsoid Plot Program for Crystal Structure Illustrations*, Rept. no. ORNL-5138 (Oak Ridge National Laboratory, Tennessee, 1971).
3. Harding, M. M., MacLennan, J. M. & Paton, R. M. *Nature* **274**, 621–623 (1978).
4. Hamilton, J. A., Sabesan, M. N., Steinrauf, L. K. & Geddes, A. *Biochem. biophys. Res. Commun.* **73**, 659–664 (1976).
5. Griffiths, D. W. & Bender, M. L. *Adv. Catalysis* **23**, 209–261 (1973).
6. Edwards, H. E. & Thomas, J. K. *Carbohydrate Res.* **65**, 173–182 (1978).

matters arising

Injection events in ocean history

THE hypothesis that sudden intermittent injections of brackish or hypersaline water from large, temporarily isolated basins into the world ocean may have so altered the existing density stratification that mass extinctions (particularly of the open ocean plankton) resulted^{1,2}, is fraught with possibilities. The mid-Cretaceous anoxic event and the Messinian evaporite event may have workable models, but we are concerned that the models for the Cretaceous-Tertiary boundary and Eocene-Oligocene event have major flaws in that both models require an isolated, fresh to brackish water Arctic Ocean. The best (and perhaps only) available evidence for Maastrichtian time indicates that the open ocean plankton of the Arctic Ocean were neither endemic nor brackish or fresh water. Deep-sea sediment core FL-437 recovered in 1969 from the Alpha Cordillera of the central Arctic Ocean³, contains a rich phytoplankton biota described originally to show specific affinities with late Cretaceous marine silicoflagellates of Russia and California⁴, and subsequently of the south Pacific and Indian Oceans⁵⁻⁷. A study of the abundant diatom flora substantiates these affinities with known normal marine species (J. A. Kitchell and D. L. Clark in preparation). The complex is similar to the middle-late Maastrichtian complexes of California and the Koryak Range of Siberia⁸. Consequently, if a brackish water injection event occurred in the latest Cretaceous, it certainly did not originate from an Arctic basin encompassing the Alpha Cordillera.

The Eocene termination similarly requires an isolated Arctic basin. Unfortunately, neither Eocene nor Oligocene deep-sea sediment has been recovered from the central Arctic, and this model cannot be tested.

However, recent tectonic interpretations suggest that the time of rotation of the Amerasia Basin was pre-late Cretaceous^{9,10}. Detailed tectonic studies in the Bering Strait region support a late Cretaceous compressional event related to opening of the North Atlantic^{11,12}. The post-Miocene benthic foraminifera and ostracode fauna of the Alpha Cordillera

exhibit both endemism and Atlantic affinities¹³. Hence, the palaeontologic data derived from the central Arctic to date, support a transition from Pacific circulation patterns to Atlantic circulation, in conjunction with the time of opening of the North Atlantic. The observed extinctions of plankton species in the North Atlantic during this time interval may consequently represent the results of competitive interactions.

D. L. CLARK
J. A. KITCHELL

Department of Geology and Geophysics,
The University of Wisconsin,
Madison, Wisconsin 53706

1. Thierstein, H. R. & Berger, W. H. *Nature* **276**, 461-466 (1978).
2. Gartner, S. & Keany, J. *Geology* **6**, 708-712 (1978).
3. Clark, D. L. *Geology* **2**, 41-44 (1974).
4. Ling, H. Y., McPherson, L. M. & Clark, D. L. *Science* **180**, 1360-1361 (1973).
5. Bukry, D. & Foster, F. H. *J. Res. U.S. Geol. Surv.* **2**, 303-310 (1974).
6. Hajos, M. *Init. Rep. DSDP Leg 29*, 913-1009 (1975).
7. Perch-Nielsen, K. *Init. Rep. DSDP Leg 29*, 677-721 (1975).
8. Strelnikova, N. I. *Beih. zur. Nova Hedw.* **53**, 311-321 (1975).
9. Sweeney, J. F., Irving, E., & Geuer, J. W. in *Arctic Geophysical Review* (ed. Sweeney, J. F.) 91-100 (Earth Phys. Br. 45, 1978).
10. Johnson, G. L. *et al. Polarforschung* (in the press).
11. Fujita, K. *J. Geol.* **86**, 159-172 (1978).
12. Patton, W. W., Jr. & Tailleux, I. L. *Bull. Geol. Soc. Am.* **88**, 1298-1304 (1977).
13. Clark, D. L. in *Paleontology and Plate Tectonics* (ed. West, R. M.) **2**, 55-76 (Mus. Sp. Pub. Biol. Geol. Milwaukee 1977).

THIERSTEIN REPLIES—There is little doubt that the Arctic Ocean was filled with marine waters at times during the late Cretaceous. In addition to the evidence cited by Clark and Kitchell of reworked late Cretaceous phytoplankton found in Plio-Pleistocene piston-cores^{1,2}, records of Maastrichtian marine microplankton succeeded by dominantly terrestrial palynomorphs are known from several basins in northern Alaska and the Canadian Arctic^{3,4}. The transition from marine to terrestrial deposition in these marginal basins, however, may be time-transgressive and could be due to local tectonic effects^{4,5}. Tectonic activity in conjunction with a worldwide regression would provide the ideal setting for a temporary isolation and subsequent flushing of the Arctic Ocean through, for instance, the M'Clure graben near the end

of the Mesozoic (refs 6, 7 and H. R. Balkwill, personal communication). Considering the short time estimated necessary for a freshening of the Arctic Ocean⁸ and the comparatively long duration of several million years of phytoplankton zones in the late Cretaceous^{9,10}, the evidence presently available from the Arctic does not preclude a temporary freshening and subsequent flushing of that basin at the very end of the Maastrichtian and Eocene.

HANS R. THIERSTEIN
WOLFGANG H. BERGER

Scripps Institution of Oceanography,
University of California, San Diego,
La Jolla, California 92093

1. Clark, D. L. *Geology* **2**, 41-44 (1974).
2. Ling, H. Y., McPherson, L. M. & Clark, D. L. *Science* **180**, 1360-1361 (1973).
3. Felix, C. J. & Burbridge, P. P. *Geosci. Man* **7**, 1-29 (1973).
4. Doerenkamp, A. *et al. Bull. Can. Petrol. Geol.* **24**, 372-417 (1976).
5. Plauchut, B. P. & Jutard, G. G. *Bull. Can. Petrol. Geol.* **24**, 321-371 (1976).
6. Daase, H. D. & Rutgers, A. T. C. *Bull. Can. Petrol. Geol.* **23**, 84-108 (1975).
7. Balkwill, H. R. & Bostin, R. M. *Geol. Soc. Am. Abstr.* **10**, 362 (1978).
8. Srivastava, S. P. *Geophys. J. R. astr. Soc.* **52**, 313-357 (1978).
9. Gartner, S. & Kearny, J. *Geology* **6**, 708-712 (1978).
10. Bukry, S. & Foster, F. H. *J. Res. U.S. Geol. Surv.* **2**, 303-310 (1974).
11. Perch-Nielsen, K. *Init. Rep. DSDP Leg 29*, 677-721 (1975).

Shelf temperatures and reef growth on the south-east Florida coast

LIGHTY *ET AL.*¹ describe a drowned Holocene *Acropora palmata* reef off the south-east Florida coast. In attempting to explain the lack of growth of this structure over the past ~7,000 yr, the authors list papers which they claim support low temperature as the principal limiting factor ("...pronounced temperature fluctuations and unusually cold bottom waters in the shallow waters off south-east Florida are effectively preventing active reef growth north of Miami."). However, none of their references apply to this point. The paper by Taylor and Stewart², for example, deals with upwelling off the north-east and central Florida coast—not

at all in the vicinity of the barrier reef. Similarly, there is no evidence that upwelling significantly alters water temperatures on the south-eastern shelf, although it may do so on the Miami Slope and Terrace³⁻⁵. While Vaughan's data^{6,7} indicate that Fowey Rocks (northern Florida Keys) occasionally becomes cooler than the rest of the Florida Reef Tract, there are indications that this is an atmospheric effect due to an open connection with easily-chilled, shallow bay waters⁸. Vaughan did not measure the temperature north of Fowey Rocks, but assumed the existence of a sharp decrease in winter shelf temperatures north of Miami.

In a more direct investigation of these reefs⁹ 27 species of scleractinians were found and it was noted that "Reduced temperatures ... inhibit coral growth further north, but appear to be insignificant in the area studied." Specifically, this meant that in an 18 month period (1969-70), the minimum bottom temperatures taken *in situ* with minimum-maximum thermometers at depths of up to 44m did not fall below 20°C. Similarly, thermistor arrays with sensors on the reef at Boca Raton and Pompano Beach, Florida, recorded no less than 22.2°C during the winters of 1968, 1969 and 1970^{10,11}. In addition, the unpublished XBT data for February 1973 referred to by Lee and Mooers³ shows the same temperature minimum at reefal depths from Fowey Rocks to Fort Lauderdale (T. Lee, personal communication). [The low datum of 17.04°C at 32 m given in the same paper seems to be a typographical error, judging from the published thermistor temperatures at the same station (T. Lee, personal communication).] These temperatures are not unexpected on such a narrow (3-5 km) continental shelf where tide and wind induced fluctuations are small compared with the influence of the warm Florida Current¹².

While it is clear that some limitation is being placed on coral growth here, there are no data which indicate that temperatures at 15-30 m off south-east Florida are much different from those at the same depth farther south, where it is generally considered that reef growth is appreciable. It would be instructive to determine accretion rates on the deep reefs of the Florida Keys where *Acropora palmata* is as absent as it is on the reefs described by Lighty *et al.*

W. M. GOLDBERG

Department of Biological Sciences,
Florida International University,
Miami, Florida 33199

1. Lighty, R. G., Macintyre, I. G. & Stuckenrath, R. *Nature* **276**, 59-60 (1978).
2. Taylor, C. B. & Stewart, Jr., H. B. *J. geophys. Res.* **64**, 33-40 (1959).
3. Lee, T. N. & Mooers, C. N. K. *Bull. mar. Sci.* **27**, 758-775 (1977).
4. Brooks, D. A. & Mooers, C. N. K. *J. geophys. Res.* **82**, 2569-2576 (1977).

5. Lee, T. N. & Mayer, D. A. *J. mar. Res.* **35**, 193-220 (1977).
6. Vaughan, T. W. *Proc. natn. Acad. Sci. U.S.A.* **2**, 95-100 (1916).
7. Vaughan, T. W. *Carnegie Inst. Wash. Pap. Dry Tortugas Lab.* **9**, 319-339 (1918).
8. Marszalek, D. S., Babashoff, Jr., G., Noel, M. R. & Worley, D. R. *Proc. 3rd int. Coral Reef Symp. Miami* **2**, 224-229 (1977).
9. Goldberg, W. M. *Bull. mar. Sci.* **23**, 465-488 (1973).
10. Lukin, L. D. (ed.) *Limitations and Effects of Waste Disposal on an Ocean Shelf* (Florida Ocean Sciences Inst., 2nd Ann. Rep., 1969).
11. Lukin, L. D. (ed.) *Limitations and Effects of Waste Disposal on an Ocean Shelf* (Florida Ocean Sciences Inst., Final Rep., 1971).
12. Lee, T. N. & McGuire, J. B. *Adv. Water pollut. Res.* **6**, 865-878 (1972).

LIGHTY ET AL. REPLY—In our report¹, we attributed the demise of the shelf-edge *Acropora palmata* (Lamarck) barrier reef off south-east Florida to periods of extensive turbidity and to pronounced decreases in water temperatures during the first flooding of large, relatively flat areas of the shelf by rising seas approximately 7,000 yr ago. We did not suggest, as Goldberg implies, that low shallow-water temperatures were the principal cause of the reef's death. Rather, these cold temperatures were one of at least two unfavourable conditions that apparently existed about 7,000 yr ago, and that today control the distribution of modern *A. palmata* reefs in south Florida².

Goldberg's comments suggest that he has misinterpreted our discussion¹ of the role of cold water temperatures in limiting reef growth during modern as well as early Holocene time on the entire south-east Florida shelf north of Miami. Goldberg has directed his attention to present-day temperatures at the shelf edge whereas we were referring to pre-existing temperatures at the shelf edge. The low water temperatures that may have contributed to the death of the shelf-edge reef would have resulted from atmospheric cooling of the large expanses of shallow water that formed when the sea initially flooded the continental shelf. Similar conditions currently exist in large shallow-water areas of south Florida, as shown by minimum water temperatures of 18° or less recorded for 23 winter months between 1954 and 1968 in northern Biscayne Bay near Miami³, and for eight winter months between 1968 and 1971 at more than 40 localities in southern Biscayne Bay⁴.

Furthermore, north of Miami present-day atmospheric cooling of shallow shelf water may be preventing *A. palmata* reef growth on the many submerged ridges in nearshore areas of the south-east Florida shelf^{5,6}. Although water depths are shallow enough for the development of *A. palmata* reefs, none were found during an extensive survey of these inner shelf ridges⁶. Along the shelf edge, however, water temperatures may be more favourable for reef growth, but the present water depths of 15-30 m over the relict barrier reef are well below depths in which an active *A. palmata* reef can exist. The hermatypic coral species found by Goldberg⁷ on the surface of the relict barrier reef were a deep reef fauna which did not

include rapid reef-forming species such as *A. palmata* and *Porites porites* (Pallas). In addition, *Acropora cervicornis* (Lamarck) and *Montastrea annularis* (Ellis and Solander) occurred in very low abundance⁷. Early Holocene radiocarbon dates for samples from close to the surface of the relict barrier reef¹ indicate that the deeper water epifaunal community has contributed no appreciable relief to this shelf-edge structure.

The temperature data to which Goldberg refers are not relevant to our discussion, nor are they as extensive as he suggests. The low values cited for "an 18 month period" do not appear in the report to which he refers. Measurements were made by thermistor probes at only one site, off Boca Raton (ref. 8, p 107), with three probes suspended vertically near the shelf edge at depths of approximately 7.5, 16.5, and 29 m, and not "with sensors on the reef." Temperature profiles along transects of the outer reef were measured only on selected dates by a portable salinometer (ref. 8, p 107, 110).

Therefore, it is evident that only minimal data exist for present-day water temperatures at the shelf edge in this area north of Miami, where depth is the most limiting factor in preventing reef growth. Temperature data are also lacking for the shallow inner shelf areas, where substrates are suitable for reef growth but lack any modern accumulation. Finally, although we cannot determine pre-existing low shallow-water temperatures at the shelf edge, we suggest¹ that they existed and that they were at least partly responsible for the death of the relict barrier reef off south-east Florida.

ROBIN G. LIGHTY

Department of Geology,
University of North Carolina,
Chapel Hill, North Carolina 27514

IAN G. MACINTYRE

Department of Paleobiology,
Smithsonian Institution
Washington, D.C. 20560

ROBERT STUCKENRATH

Radiation Biology Laboratory,
Smithsonian Institution
Washington, D.C. 20560

1. Lighty, R. G., Macintyre, I. G. & Stuckenrath, R. *Nature* **276**, 59-60 (1978).
2. Ginsburg, R. N. & Shinn, E. A. *Bull. Am. Ass. petrol. Geol.* **48**, 527 (1964).
3. Moore, H. B. *Miami Sea Temperatures and Salinities 1-8* (Institute of Marine Sciences, Univ. Miami, Tech. Report ML 69039, 1969).
4. Roessler, M. A. & Tabb, D. C. *Studies of Effects of Thermal Pollution in Biscayne Bay, Florida 122-129* US Environmental Protection Agency Report 660/3-74-014, (1974).
5. Duane, D. B. & Meisburger, E. P. in *Geomorphology and Sediments of the Nearshore Continental Shelf, Miami to Palm Beach*, 8-14 (U.S. Army Corps of Engineers Coastal Eng. Res. Center Tech. Memo 29, Washington, 1969).
6. Raymond, B. thesis, Florida State Univ. (1972).
7. Goldberg, W. M. *Bull. mar. Sci.* **23**, 465-488 (1973).
8. Lukin, L. D. (Ed.) *Limitations and Effects of Waste Disposal on an Ocean Shelf* 105-137 (Florida Ocean Science Inst., Final Rep. 1971).

reviews

Survival value of migration

J. R. Krebs

The Evolutionary Ecology of Animal Migration. By R. R. Baker. Pp. 1024. (Hodder and Stoughton: London, 1978.) £35.

MIGRATION is a subject which crops up at every major ethological and ecological conference. Orientation and navigation mechanisms of migratory birds, seasonal migrations of vast herds of African ungulates, and periodic emigrations of microtine rodents are among many migration phenomena which never fail to fascinate zoologists. It is surprising in view of the pervasive interest in migration in one form or another that Baker's book is the only modern attempt to deal with the subject in a comprehensive manner. The book covers a taxonomic range from earthworms to man. It boldly attempts to present a general theory of the survival value of migration, and integrate it with what is known about physiological mechanisms.

It is an awe-inspiring work; impressive in its size (over 1000 pages long, even the figure captions sometimes spread over three pages), the illustrations are excellent, and the detail with which many examples are described is remarkable. At the same time one cannot help feeling that the size and thus inaccessibility of the book will discourage many people from giving it the attention it deserves. Baker could have presented the essence of his argument in a much shorter book and reached a much bigger audience.

Baker starts with a 30 page discussion of the definition of migration. Ornithologists use the term specifically to refer to regular, seasonal, long distance, return movements. These are characteristic, for example, of the many birds breeding in temperate zones and wintering in warmer areas. Insects rather rarely make return journeys (at least within a single individual's lifetime) so that entomologists do not characterise migration by the criterion of return to a starting point. In fact, Baker argues, all the separate features which can be used to characterise migration such as distance, return and periodicity vary along a continuum. The upshot is that migration is defined by Baker in the most general way possible: "the act of moving from one spatial unit to another". This allows Baker to include

both an aphid flitting a few centimetres from one leaf to the next, and an Arctic Tern making a 16,000 km pole to pole trip under the rubric of migration. Although there are obvious disadvantages to such a wide definition (a branch falling off a tree "migrates" in Baker's terms), Baker argues that there may be general rules which apply, as he puts it, "to the black bean aphid, the blue whale and the barnacle".

Baker's general cost-benefit model of migration is developed in the next 50 pages of the book. The essence of the argument is that it is in theory possible to calculate for an individual at any moment in time the expected fitness resulting from staying put, and the expected fitness resulting from a decision to move to a new place. When the expected gain for moving is higher than that for staying, migration will be favoured. Baker derives the basic model of an optimal migration threshold, and then goes on to discuss various ramifications. These include stochastic variation in habitat quality, the effects of competition within a habitat on migration thresholds, and exploratory migration as a method of acquiring information about expected pay-offs. Most of the theoretical arguments are qualitative, and one sometimes feels that a rigorous derivation would make the point more convincing. An example is the statement on p72 that "unpredictable fluctuations in E (the expected fitness consequent upon migration) . . . are more likely to lead to a . . . distribution of staytimes . . . corresponding to the distribution of frequency of different values of E ".

Baker's model of migration is essentially the same as the well known "marginal value theorem" published several years ago by E. L. Charnov, G. A. Parker, and others. Baker's application of the idea to migration in general is perhaps somewhat novel, but to some extent the edge is taken off the theoretical section of the book by the fact that the ideas have already been developed by others.

The remaining 900 pages of the book are entitled "evaluation of the model: examples and analysis". The organisation is mainly taxonomic. As there are virtually no data which can be used to test the model in a rigorous way,

by far and away the bulk of the "evaluation" is in terms of qualitative evidence and *post hoc* interpretation. One quantitative test of the model which Baker describes is the outstanding work of G. A. Parker on dungflies, and this occupies only two or three pages. I feel it could have been given more extensive coverage as it is so central to Baker's theme. Although the literature seems to be well covered, there are some obvious gaps. For example, the section on movement paths in chapter 18 does not refer to J. Paloheimo's theoretical analysis of optimal search paths, or J. N. M. Smith's refined experimental analysis of bird movement tracks. The long section on bird navigation and pigeon-homing makes no reference to the Italian work on scent maps, Keeton's work on release site bias, or Gwinner's re-appraisal of the star map work of Sauer. One general area which did not seem to be covered is the genetics of migration. The theoretical arguments assume that there is genetic variance for migration thresholds, but the work on genetic correlates of emigration in small mammals, for example, is not included.

However, these minor omissions are more than offset by the great diversity of facts and insights in the book. To mention just one example, the discussion of home range and territoriality (pp384-410) places a useful emphasis on the exploratory behaviour of non-territorial individuals. In part, the discussion centres around conditions which will favour a non-territory holder attempting to displace a resident instead of migrating to try and find an empty space elsewhere. Looking at home range defence from the resident's point of view, Baker argues that one of the most general benefits derived from exclusion of others is an increase in the accuracy with which resources in different parts of the home area can be assessed and ranked. If transient individuals are permitted to exploit these resources, then the value of exploration and ranking by the resident is reduced. As with so many of the ideas in the book, there is plenty to stimulate thought and discussion, but no hard evidence.

Baker's discussions of migration in modern industrial human societies is

certain to arouse controversy. He takes the extreme view that our present-day migration patterns are directly determined by natural selection. He also argues that natural selection is still acting in a similar fashion to that experienced by our hunter-gatherer ancestors. The argument is roughly as follows. In ancestral hunter-gatherers, newly formed pairs migrated away from their natal groups to other parts of the home range explored during adolescence. Pairs tended to settle in areas where their expected reproductive success (limited mainly by food supply) was highest. Males tended to explore larger ranges than females, partly because they could increase their reproductive success by incidental copulations on the way. Baker goes on to apply these speculative conclusions to industrial man: "I feel some justification in claiming that the migration pattern shown by modern industrial humans is that evolved during the 10–20 million years of hunting-gathering. All that remains is to consider whether the

selection that is acting at the present time on modern industrial humans is tending to perpetuate or change the pattern." His highly controversial conclusion is that our present-day migration patterns are genetically controlled, hormonally mediated ("part of the physiological syndrome that transforms an individual to physical maturity is a decreasing migration threshold") behaviours. There is likely to be much argument against this conclusion. Even if one accepts in principle that genetic selection can still influence migration behaviour in man (a view for which no evidence is given), it is not at all clear to what extent factors such as extramarital copulation and protein intake alter inclusive fitness.

Although one may not agree with all of Baker's views, the book is thought provoking, and it will be a pity if it is not widely read because of its size and price. □

John R. Krebs is Lecturer in Zoology at the Edward Grey Institute of Field Ornithology, Oxford, UK.

Inherited behaviour and reactions to drugs

Drugs and the Inheritance of Behavior: A Survey of Comparative Psychopharmacogenetics. By P. L. Broadhurst. Pp. 206. (Plenum: New York and London, 1978.) £12.28.

P. L. BROADHURST has written a short but surprisingly complete book in the area of the effects of inherited behaviour on reactions to drugs. The title that he has taken was not his first choice and it is somewhat misleading. As everyone familiar with the field will know, this is a relatively small and new discipline which studies the effect of genotype on an organism's reaction to drugs. Almost all of the work discussed comes from rat and mouse experiments. The title implies that the drugs are altering the inheritance of behaviour, but of course the opposite is the case; it is the influence of inheritance on the response to drugs that is being studied. Dr Broadhurst would have preferred the title *Psychopharmacogenetics* but that title was pre-empted by B. E. Eleftheriou with an edited book (Plenum: New York, 1975). Others, such as McClearn, favour the name "Behavioural Pharmacogenetics" for the area.

Whatever the name, it is a field that has been entered on the one hand by behavioural geneticists often with insufficient knowledge of pharmacology, on the other by pharmacologists with insufficient knowledge of genetics or by psychologists with insufficient knowl-

edge of both of the other two disciplines. Dr Broadhurst's book is replete with examples of the resultant chaos. He is properly critical of the investigators who have perpetrated some of the scientific monstrosities which often seem to have been undertaken with bad design, poor or non-existent controls, poorly analysed and then published repeatedly.

There are, however, numerous examples of well designed and well executed studies. Students and those

Aflatoxins

Aflatoxins: Chemical and Biological Aspects. By J. G. Heathcote and J. R. Hibbert. Pp. 212. (Elsevier: Amsterdam and New York, 1978.) Dfl. 120; \$53.50.

MANY processed food components, and the foods themselves, are potential substrates for moulding by mycotoxin-producing fungi and so it is no small wonder that aflatoxin B, being the most potent hepatocarcinogen in the rat, and related aflatoxins should be a subject of much biological and chemical research.

After opening chapters on the discovery of aflatoxins and their production under natural conditions and in the laboratory the chemistry of these substances is reviewed and their assay by physicochemical techniques and bioassay are described.

Thereafter, sufficient pathology is presented for the non-medical reader to appreciate the general nature of the effects without being blinded by technical jargon. The distinction between acute toxicity and carcinogenicity is made clear.

just entering the field would do well to read this book and take heed of the lessons that it contains. It gives examples of the use of the most popular and powerful genetic techniques including selective breeding, sex differences, strain differences, diallel crosses and recombinant inbreeding.

For the novice to the field the precautions and controls that are outlined seem excessive and unnecessary; however, the experiments usually require months if not years of study and are often exceedingly expensive. The techniques are extremely powerful but much time and money can be wasted unless careful planning is done before rather than during or after the experiment.

The book has the decided advantage of being written by one individual with a consistent style and approach. It also has the advantage over edited books of being less self-serving and providing a broad overview rather than a series of narrow reports. In his desire to be complete, Dr Broadhurst has done a remarkable job, but therein also lies one of the criticisms of the book. In his zeal he has often used information contained in abstracts. Generally this is done because the data has not appeared in an edited journal. Thus, some of the conclusions reached by Dr Broadhurst rest on some rather thin ice.

Richard A. Deitrich

Richard A. Deitrich is Professor and Vice-Chairman of Pharmacology at the University of Colorado School of Medicine, and Research Associate of the Institute for Behavioral Genetics, Denver, Colorado.

The chapters on biochemistry and metabolism perhaps constitute the most intellectually stimulating feature of this book and the authors, who have worked in the field, have not hesitated to make certain value judgements. They are cautious concerning the significance of the elusive derivative aflatoxin B₁-2-3 epoxide in acute toxicity and offer hemiacetal derivatives of aflatoxin as candidates for the role of carcinogen.

The book concludes with a well illustrated review of the biosynthesis of the aflatoxins and their interrelationships, followed by a chapter on control of aflatoxin contamination and current attempts at detoxification of foods.

The style of the text is clear and readable and typographical errors are rare. Those concerned with food science should read this book. I also suspect that the general reader could catch from it the fascination of aflatoxins and their mode of action.

P. G. Mantle

P. G. Mantle is Senior Tutor in Biochemistry at Imperial College, University of London, UK.

Chemisorption of gases on metals

Chemisorption of Gases on Metals. By F. C. Tompkins. Pp. 370. (Academic: London, New York and San Francisco, 1978.) £16.80.

MAJOR advances in our knowledge and understanding of the chemisorption of gases on metals have occurred during the past decade. The elemental composition, crystallography, and the electronic and vibrational properties of surface layers on defined single crystal faces are now measurable, and a wealth of detailed information is available in the research literature. But for the advanced undergraduate or beginning postgraduate student there has been a serious lack of comprehensive introductory texts in which the theoretical bases and experimental methods of the newer physical techniques are clearly explained and in which the contributions of these techniques are placed in perspective with the traditional methods of surface study.

Professor Tompkins' book is intended to fill this gap. Its plan is excellent. The early chapters deal with the terminology of adsorption, methods of preparing clean surfaces, the kinetics of adsorption and desorption, isotherms and heats of adsorption, and the thermodynamics and statistical thermodynamics of adsorption. These chapters provide a natural link with earlier texts. They are followed by outlines of the electronic theory of metals and metal surfaces. The remaining half of the book is composed of short chapters dealing with work functions, magnetic susceptibility, electrical conductivity, low energy electron diffraction, infrared spectroscopy, field emission and field ion microscopy, Auger electron spectroscopy, photoelectron spectroscopy, ion neutralisation spectroscopy, electron energy loss and appearance potential spectroscopies, and a short review of some experimental results for adsorbates on tungsten.

The book conveys quite well the interplay of theory and experiment in modern surface science, and gives due emphasis to the methods which have proven most fruitful. However, it is neither as comprehensive nor as up to date as one might have expected. Ion-scattering and secondary ion mass spectrometry are completely omitted, as is ellipsometry, and the only reference to high resolution electron energy loss spectroscopy is the early paper of Propst and Piper (1967). Although published towards the end of 1978, the book contains very few ref-

erences dated after 1975. Except for the treatment of ultraviolet photoelectron spectroscopy, it is not appreciably more up to date than Wedler's *Chemisorption: An Experimental Approach* (Butterworths, 1976) and it is less comprehensive in its coverage of experimental methods. More emphasis is placed by Tompkins on the theoretical background, particularly the electronic properties of metals and surfaces. However, the level of the theory is sometimes too ambitious for its proper development in an intro-

ductory book of this scope and the student will often be baffled unless he pursues the references.

Despite these shortcomings the book is a welcome addition to the literature on chemisorption. The student will find it a useful guide to an increasingly interdisciplinary field of physics and chemistry.

J. Pritchard

J. Pritchard is Professor of Physical Chemistry at Queen Mary College, University of London, UK.

Account of the bumblebee

The Life of the Bumblebee. By D. V. Alford. Pp. 80. (Davis-Poynter: London, 1978.) £3.95.

THIS is a popularised and much abbreviated version of D. V. Alford's *Bumblebees* which was published by Davis-Poynter in 1975. To help the general reader Alford begins by reviewing the features of insects in general, and bumblebees in particular. He then follows the usual and successful formula of tracing the biology of the colony from its initiation in spring, its growth to maturity, and the production of males and queens. He goes on to discuss further topics (including temperature regulation, foraging, orientation, mating and hibernation) and refers to most of the more important recent findings as well as the better

known and established aspects of bumblebee biology.

The final chapter entitled "Bumblebees and Man" is most valuable in drawing the reader's attention to the importance of bumblebees as pollinators of our crops, and the need whenever possible to protect them from the hazards of insecticides, to replace sources of forage destroyed by herbicides and intensive cultivation, and to provide sites suitable for nesting and hibernation. Appendices describe the 25 British species of bumblebees. There are eight pages of good black and white photographs.

In general Alford is to be congratulated in avoiding placing undue emphasis on his own particular interests and in presenting a well balanced accurate and readable account of the bumblebee.

John B. Free

John B. Free is at Rothamsted Experimental Station, Harpenden, UK.

Biological identification methods

Biological Identification: The Principles and Practice of Identification Methods in Biology. By R. J. Pankhurst. Pp. 104. (Edward Arnold: London, 1978.) Paperback £3.20.

SYSTEMATICS as an approach to the teaching of the biologies was becoming unpopular in the 1960s. Memory-taxing, as then taught, it carried a flavour of the elite natural history museums and culture collections: the spirit of Robbins was moving university teachers towards molecular biology and genetics which could, in some fashion, be assimilated by the many.

Meanwhile, computers were making possible numerical, multivariate and probabilistic treatments of biological classification and identification. It has taken over a decade for this to penetrate university teaching and by now there is a shortage of young systematists with an interest in the identification craft.

R. J. Pankhurst's compact book on biological identification methods begins with a splendidly lucid section on the use and construction of keys, taking British species of the genus *Epilobium* as a running example. If the rest were as easy to understand, one could begin to see a very satisfying second-year biology degree module in general systematics drawing on the second edition of C. Jeffrey's *Biological Nomenclature* (Edward Arnold: London, 1977) for nomenclature, P. H. A. Sneath and R. R. Sokal's *Numerical Taxonomy* (Freeman: Reading, 1973) for classification, and Pankhurst for identification. Lecturers should buy their copy well ahead of the students and settle down to some supplementary reading (the short bibliography is well chosen): the section on probabilistic methods drawing on Bayes' theorem, for a start, will need cracking by a worked-out example, that given (Fig. 16) being just a section of a computer print-out, understandable only by the initiated.

Monothetic and polythetic methods, explained on page 3, would have been an academically sound way of arranging the core of the work. Instead, the

central chapters focus respectively on "conventional" and "automatic" methods: the latter "may or must involve calculations . . . so that the use of a computer is usual". Though it perhaps looks sensibly down-to-earth, this is a comparison not of identification methods but of the ways in which we find it convenient to go about them, and it leads to some confusion. Diagnostic keys are "conventional", are they not? But, wait, we can programme computers to construct keys, so that must be "automatic". Matching methods, now these are certainly "automatic": botheration, they have to go in the "conventional" section as well because we can identify by comparison without calculation. And so on, until it must

Ecological methods

Ecological Methods, with Particular Reference to the Study of Insect Populations. Second edition. By T. R. E. Southwood. Pp. 524. (Chapman and Hall: London, 1978.) £10.

It is a mark of the uniqueness of this book that in the 12 years since the first edition no competitor volume has emerged that covers such a wide field. Professor Southwood is just as much at home discussing the idiosyncracies of insect trap construction as he is comparing indices of diversity.

The book is well described by its title, and like many compendia will be most useful to the expert or semi-expert. The sort of methods we choose are influenced by our current understanding of the ecological world around us. Nevertheless, in many asides the author does try to redress the balance, pointing out the relative merits of the different techniques, what they can and cannot do. It is here that the book will also be useful to the beginner.

In the past 12 years there has been a dramatic but allometric growth of ecology in two quite different areas. On the one hand the subject has im-

dawn on the undergraduate or advanced school student that it is identification logic to which he should address himself, not the fuss about computing and programming language (leading, incidentally, to a strange Fig. 18 which actually suggests that "on-line" is a method of biological identification).

It will be observed that the reviewer has been stimulated. So will be the many others who buy this interesting, thought-provoking volume; they, too, will have to work hard.

J. E. Smith

J. E. Smith is Professor of Microbiology at the University of Surrey, Guildford, UK.

pinged more on public consciousness; on the other, the rapid development of computers and electronic calculators has stimulated much mathematical modelling and analysis. The result is that the scope of ecology and of this second edition is much wider than ever before.

Every chapter has been updated and new sections added, especially on dispersion, capture-recapture methods, components of predator-prey interaction, and dispersal. In particular, recent growth in the methods of construction of life-budgets, systems analysis, descriptions of diversity and species-packing models now allow Professor Southwood to crystallise out a conceptual framework for discussion which was not possible before.

Theoretical astrophysics

Theoretical Principles in Astrophysics and Relativity. Edited by N. R. Lebovitz, W. H. Reid and P. O. Vandervoort. Pp. 257. (University of Chicago Press: Chicago and London, 1978.) £16.10; \$23.

Theoretical Principles in Astrophysics and Relativity offers yet another collection of papers from a conference, this one held in honour of Professor S. Chandrasekhar's 65th birthday in 1975. I found it a distinct improvement on the usual hotch-potch of disparate lectures re-hashed for the wider public with a minimum of editorial tinkering. Firstly, the book is better produced than most conference reports (which usually resemble a wedge of preprints glued together)—which probably accounts for the immense delay in publication. More importantly, however, the articles are mainly of a review rather than research variety; this is fortunate as nearly four years have

A reviewer must also criticise. The greatly expanded field of the subject (1,000 new references, with few deletions from the first edition) is covered in the space of 130 extra pages. The pace of the discussion is therefore fast and, occasionally, confusing. There are omissions (for example, the log-normal distribution is mentioned twice, on pp33 and 424, without its basic model) and repetitions (for example, equations 2.26, 2.27, 2.29 and 2.30 re-appear as equations 2.37, 2.38, 2.41 and 2.42); and occasionally parameters are ill-defined (for example, I_m of equation 9.28 and S_T of 13.1) or, together with variables, given different symbols in the text and figures (for example, q and Q in Table 9.1 and equation 9.6, N_E , N_R and P_E , P_R on p382 and Fig. 10.8). Generations of students will now continue to write, incorrectly, of density dependance (Fig. 10.7, in contrast to Fig. 89 of the first edition) and of the Shannon-Weaver rather than Shannon-Wiener function (p426). There are also several small, typographical errors.

None of this, however, detracts from the significant contribution of this volume. The author and publishers are to be congratulated on producing an expanded second edition at less than the price of the final printing of the first. By far the best value on any ecologist's bookshelf.

D. J. Rogers

D. J. Rogers is Lecturer in Entomology at the University of Oxford, UK.

elapsed since the lectures were prepared.

I found the paper on stellar dynamics by Contopoulos especially lucid, and was surprised to see a very formal mathematical paper by Mullikin on the very physical subject of radiative transfer. There is a nice review by Thorne on various phenomena in astrophysics in which general relativity plays a central role, looking all the more relevant with the recent probable discovery of gravitational radiation. The article on singularities in spacetime by Penrose is masterly and thought provoking.

The range of topics covered in the conference was fairly wide, and the articles necessarily somewhat superficial. The book does have, however, a rather more permanent quality about it than comparable proceedings, and is certainly good introductory reading for a host of topics in astrophysics and cosmology. At £16.10, it is not overpriced for such a well-produced format.

P. C. W. Davies

P. C. W. Davies is Lecturer in the Department of Mathematics at King's College, University of London, UK.

● The US edition of *The Climatic Threat* by John Gribbin (for review, see *Nature*, 272, 788; 1978) has been published as *What's Wrong with Our Weather?* by Scribners: New York (\$9.95).

○ The price of *Principles of Chemical Kinetics* (for review, see *Nature*, 278, 194; 1979) was incorrectly quoted. This should be £12.65.

● The US edition of *The Visible College* by Gary Werskey (for review, see *Nature*, 276, 136; 1978) will be published in May by Holt, Rinehart and Winston.